Department of Food Hygiene and Environmental Health

Faculty of Veterinary Medicine

University of Helsinki

Helsinki, Finland

The origin and genetic basis of nonulosonic acid biosynthesis in *Campylobacter coli* lipooligosaccharides

Alejandra Culebro Escandón y Pérez

Academic dissertation

To be presented, with the permission of the Faculty of Veterinary Medicine of the University of Helsinki, for public examination in sali 104 (ls B7), Metsätieteiden talo, Latokartanonkaari 7, Helsinki, on June 1st 2018, at 12 noon.

Helsinki 2018

Supervising Professor

Professor Maria Fredriksson-Ahomaa, DVM, PhD
Department of Food Hygiene and Environmental Health
Faculty of Veterinary Medicine
University of Helsinki
Helsinki, Finland

Supervised by

Assistant professor Mirko Rossi, DVM, PhD
Department of Food Hygiene and Environmental Health
Faculty of Veterinary Medicine
University of Helsinki
Helsinki, Finland

Reviewed by

Professor Victoria Korolik, PhD
Institute for Glycomics
Griffith University
Southport, Australia

Senior lecturer Arnoud van Vliet, PhD
Department of Pathology and Infectious Diseases
School of Veterinary Medicine
Faculty of Health and Medical Sciences
University of Surrey
Guildford, UK

Opponent

Professor Christina Schäffer, Dipl.-Ing. Dr.nat.techn
Department of NanoBiotechnology
NanoGlycobiology unit
Universität für Bodenkultur (BOKU)
Vienna, Austria.

Dissertationes Schola Doctoralis Scientiae Circumiectalis, Alimentarie, Biologicae
ISBN 978-951-51-4309-9 (paperback)
ISBN 978-951-51-4310-5 (PDF)
Unigrafia, Helsinki 2018

Table of Contents

ABSTRACT

Campylobacter jejuni and *Campylobacter coli* are the most common cause of bacterial gastroenteritis, affecting approximately 96 million people around the world every year. Campylobacteriosis is characterized by diarrhoea, abrupt abdominal pain, and fever. Generally, the disease is self-limiting and resolves within a few days. Occasionally, neurological symptoms manifest clinically three weeks post-infection, indicating the onset of Guillain-Barré syndrome (GBS). GBS is a polyradiculoneuropathy characterized by an acute progressive and symmetrical motor weakness of the extremities with varying degrees of areflexia.

GBS has been established as a post-infectious sequelae, whose principal aetiological agent is *C. jejuni.* The *C. jejuni*-GBS link was the first confirmed case of human autoimmune disease caused by ganglioside mimicry. *C. jejuni* lipooligosaccharides (LOS) have been established as the main structure responsible for ganglioside mimicry.

Studies on the genetic basis of ganglioside-like LOS expression in *C. jejuni* have identified the genes associated with the addition of sialic acid (Neu5Ac) to the LOS chain; *neuA, neuB, neuC,* and a gene encoding a sialyltransferase from the glycosyltransferase CAZy family 42 (GT-42). These genes are found within the LOS biosynthesis locus of *C. jejuni.* The existence of *neuA, neuB, neuC,* and GT-42 genes in *C. coli* is unknown, despite the fact that this species has also been isolated from GBS patients.

The primary aim of this study was to investigate the diversity of the *C. coli* LOS biosynthesis locus with particular interest in identifying genetic features associated with the synthesis and transfer of Neu5Ac. *C. coli* was found to possess a more diverse LOS biosynthesis locus than previously thought. A total of 27 different LOS locus classes containing a GT-42 encoding gene were identified. Among these, 16 are potentially able to synthesize sialylated LOS structures. Interestingly, several LOS locus classes resemble those of *C. jejuni* involved in ganglioside mimicry. Thus, bacterial factors implicated in GBS aetiology can cross species barriers. Also, *C. coli* has a larger GT-42 enzyme repertoire than *C. jejuni*. Two of the most common GT-42 encoding genes in *C. coli* were found to be associated to the presence of nonulosonic acid in LOS structures. Marked differences in diversity in the LOS locus were observed between *C. coli* clades, suggesting a potential role of this structures in niche adaptation. The importance of LOS to *C. coli* ecology and host-pathogen interaction remains to be explored.

ACKNOWLEDGEMENTS

This thesis was carried out at the Department of Food Hygiene and Environmental Health during 2014-2018. The studies included in this thesis were funded by the University of Helsinki three years research grant (313/51/2013), the University of Helsinki Microbiology and Biotechnology graduate program, the Walter Ehrström Foundation travel grant, and the Academy of Finland through the Centre of Excellence on Microbiological Food Safety no. 141140.

I am grateful to my supervisor, Assistant professor Mirko Rossi, with whom we adventured into the unknown, to us, field of microbial glycobiology. His understanding and supportive nature were key in dealing with the up and downs of this thesis. Professor Maria Fredriksson-Ahomaa is acknowledged for helping with the ever changing and lengthy bureaucracy of the thesis submission process. Johanna Laine is acknowledged for all the help provided in the administrative area.

Professors Jukka Finne and Mikael Skurnik are thanked for providing guidance and support as my thesis committee. I am also grateful to Astrid Skarp-de Haan, PhD, Thomas Schott, PhD, Joana Revez, PhD, Professor Elke Schweda, Professor Emerita Marja-Liisa Hänninen, Ben Pascoe, PhD, Yasmin Friedmann, PhD, Matthew Hitchings, PhD, Jacek Stupak, MSc, Professor Samuel Sheppard, Jianjun Li, PhD, Miguel Machado, MSc, João Carriço, PhD, and Michel Gilbert, PhD, for being co-authors in the studies included in this thesis.

I am particularly thankful to Professor Warren Wakarchuk for being an inspiring scientist who introduced me to the sweetness of microbial glycans and for teaching me how to sustainably carry out research for years.

I thank past and present members in the research group, Ann-Katrin Llarena, PhD, Satu Olkkola, PhD, Sara Kovanen, PhD, Rauni Kivistö, PhD, as well as other personnel in the department. In addition, Urszula Hirvi is warmly thanked for her support in the growth of bacterial cell mass for NMR (results that are yet to see the light), and for helping me to develop my Finnish skills.

I also want to thank my dearest friends Zhen Zeng and Junko Sugano for all the weekend activities and scientific talks that kept me focused on my goals while being in touch with life outside the lab. I am deeply grateful to my father who understood my desire to broaden my horizons, and supported me unconditionally. Our sporadic, but lengthy scientific conversations helped me to grow into a well-rounded scientist. To my mother, whose sacrifices and determination were essential in helping me reach my dreams, I have no words to thank her for always supporting even my most outrageous ideas. To Erik, who brighten up my life, I am deeply thankful for always being there encouraging me to keep on accomplishing my/our goals.

Finally, I want to dedicate this thesis to those who radically changed my life: Ville Valo, Mikko Paananen, Mikko Lindström, Mika Karppinen, Janne Puurtinen, Lauri Ylönen, Eero Heinonen, Pauli Rantasalmi, Aki Hakala, Eicca Toppinen, Perttu Kivilaakso, Paavo Lötjönen, Mikko Sirén, Tarja Turunen, Tuomas Holopainen, Emppu Vuorinen, Jukka Nevalainen, and Marco Hietala. Their music helped me to keep on dreaming of the land of the midnight sun even in the darkest moments of my life, and kept me company during the thousands of hours spent in the laboratory.

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications referred in the text by their roman numerals:

I. Skarp-de Haan, C. A. Culebro, T. Schott, J. Revez, E. Schweda, M. Hänninen, and M. Rossi. 2014. Comparative genomics of unintrogressed *Campylobacter coli* clades 2 and 3. BMC Genomics. 15:129. DOI: 10.1186/1471-2164-15-129

II. Culebro, A*., J. Revez, B. Pascoe, Y. Friedmann, M. D. Hitchings, J. Stupak, S. K. Sheppard, J. Li, and M. Rossi. 2016. Large sequence diversity within the biosynthesis locus and common biochemical features of *Campylobacter coli* lipooligosaccharides. J Bacteriol. 198:2829. Publication DOI: 10.1128/JB.00347-16. Pre-print available at bioRxiv DOI: 10.1101/050328

III. Culebro, A*, M.P. Machado, J.A. Carriço, and M. Rossi. Origin, evolution, and distribution of the molecular machinery for biosynthesis of sialylated lipooligosaccharide structures in *Campylobacter coli.* Pre-print available at bioRxiv DOI: 10.1101/225771

IV. Culebro A*, M. Gilbert, J. Stupak, J. Li, W. Wakarchuk, M. Rossi. Genetics behind the biosynthesis of nonulosonic acid containing lipooligosaccharides in *Campylobacter coli.* Pre-print available at bioRxiv DOI: 10.1101/254235

Additionally, previously unpublished material is presented.

*Corresponding author

ABBREVIATIONS

AA	auxiliary activity
Bac	2,4-diacetamido-2,4,6-trideoxyglucopyranose, bacillosamine
BLAST	Basic Local Alignment Search Tool
BBH	Best Blast Hit
BDP	boron-dipyrromethene
BIBH	Best Informative Blast Hit
BSR	Blastp Score Ratio
CAMPs	Cationic Antimicrobial Peptides
CAT	Chloramphenicol Acetyltransferase Cassette
CAZyDB	Carbohydrate-Active enZymes database
CAZymes	Carbohydrate-Active enZymes
CC	Clonal Complex
CE	Carbohydrate Esterase
CD14	Cluster of Differentiation 14
CPS	Capsule
DNA	Deoxyribonucleic Acid
EA-OTLC-MS	Electrophoresis-Assisted Open Tubular Liquid Chromatography–Electrospray Mass Spectrometry
EPEC	Enteropathogenic *E. coli*
ETEC	Enterotoxigenic *E. coli*
FCHASE	fluorescein
Gal	Galactose
GBS	Guillain-Barré Syndrome
GDP	Guanine Diphosphate
GlcN	D-Glucosamine
GlcNAc	N-Acetylglucosamine
GlcN3N	2,3-Diamino-2,3-Dideoxy-D-Glucose
GO	Gene Ontology
GH	Glycoside Hydrolases
GT	Glycosyltransferase
D,D-Hep	D-Glycero-D-Manno-Heptopyranose
L,D-Hep	L-Glycero-D-Manno-Heptopyranose
HGT	Horizontal Gene Transfer
HMM	Hidden Markov Model
HPAEC-PAD	High Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection
HT	Homopolymeric Tracts
IPTG	isoPropyl-β-D-Thiogalactopyranoside
KDO	3-Deoxy-D-Manno-Octulosonic Acid
Leg5Ac7Ac	Legionaminic Acid

LOS	Lipooligosaccharide
LPS	Lipopolysaccharide
ManNAc	*N*-Acetylmannosamine
MCL	Markov Cluster Algorithm
MFS	Miller Fisher Syndrome
MLST	Multilocus Sequence Typing
NBA	Nutrient Blood Agar
Neu5Ac	*N*-Acetylneuraminic Acid
NMR	Nuclear Magnetic Resonance
NCBI	National Center for Biotechnology Information
OM	Outer Membrane
ORF	Open Reading Frame
PEP	Phosphoenolpyruvate
PEtn	Phosphorylethanolamine
PPEtn	Pyrophosphoethanolamine
PL	Polysaccharide Lyase
PPR	Peptide Pattern Recognition
Pse5Ac7Ac	Pseudaminic Acid
PSI-BLAST	Position-specific Iterated BLAST
Qui3N	3-Amino-3,6-Dideoxy-D-Glucose
Qui3NAc	3-Acetamido-3,6-Dideoxy-D-Glucose
RAST	Rapid Annotation Using Subsystems Technology
RDP	Ribosomal Database Project
Rha	Rhamnose
SSR	Simple Sequence Repeat
STEC	Shiga Toxin-producing *E. coli*
TLR4-MD2	Toll-Like Receptor 4-Myeloid Differentiation factor 2
TNFα	Tumour Necrosis Factor alpha
UDP	Uridine Diphosphate
WGS	Whole Genome Sequencing
WHO	World Health Organization
YLDs	Years Lived with Disability

1 INTRODUCTION

In nature, there are four major classes of macromolecules: lipids, DNA, proteins, and carbohydrates. In recent years, the importance of sugars and glycoconjugates for the biology of diverse organisms has come to light. Glycobiology, a term devised in 1988 by Raymond Dwek, is an expanding field involving the study of carbohydrates in a biological setting (1). Among living organisms, bacteria synthesize a wide range of complex sugars which is about six times more diverse than what is found in eukaryotes. Since sugars are located on surface structures, they frequently play a role in cell-host interaction and serve as a first line of defence against the harsh environment. In fact, pathogens which lack classical virulence factors may heavily rely on glycoconjugate structures to modulate the host immune response (2). Such is the case of the Gram-negative, microaerobic, nonfermenter, nonsporeformer *Campylobacter* spp. *Campylobacter jejuni* and *Campylobacter coli* are common cause of bacterial gastroenteritis in the world, accounting for approximately 96 million cases every year (3). Given the fundamental role of glycans in the biology and pathogenesis of *C. jejuni* and *C. coli* (4, 5), approximately 10% of the genome is estimated to encode glycosyltransferases and other enzymes involved in the synthesis and modification of sugar nucleotides (6). The lipooligosaccharide (LOS) biosynthesis locus is one of the four major glycan biosynthesis loci. As the name suggests, genes within this locus are involved in the biosynthesis of LOS structures. Research on the genetic diversity of the LOS biosynthesis locus has been driven by the ability of *C. jejuni* to express LOS structures resembling human glycans, particularly gangliosides (7-10). Occasionally, infection with a *C. jejuni* strain expressing ganglioside-like LOS may trigger the development an autoimmune disease known as Guillian-Barré syndrome (GBS) (11). According to estimates, between 28-45% of GBS cases are due to *Campylobacter* infection (3). Genes involved in synthesis (*neuA, neuB1,* and *neuC*) and transfer (glycosyltransferases from the CAzy family 42 (GT-42) (12)) of sialic acid (Neu5Ac) are essential for the expression of ganglioside mimics in *C. jejuni*. So far, these genes have been found in the *C. jejuni* LOS biosynthesis locus genetic classes A, B, C, M, R, or V. All studied *C. jejuni* strains carrying one of these classes have been shown to express ganglioside mimics, among other sialylated LOS structures. In few instances, *C. coli* has also been isolated from GBS patients (13-16). However, the overrepresentation of *C. jejuni* and the deceptive overlap that this two species have in ecology, epidemiology, and pathogenesis has resulted in *C. coli* being vastly neglected. No classes containing genes involved in LOS sialylation were identified in the single study addressing *C. coli* LOS locus diversity (17). Moreover, excluding a single *C. coli* strain (16), no LOS sialylation has been found in the few studied *C. coli* strains (18-22). Thus, the scarce research conducted so far, is insufficient to determine whether a link between *C. coli* and GBS exists. In this study the origin, evolution, and distribution of the molecular machinery for biosynthesis of sialylated LOS structures in *Campylobacter coli* is addressed. Additionally, insights into the function of *cstIV* and *cstV,* the most common *C. coli* LOS associated GT-42 genes, are provided.

2 REVIEW OF THE LITERATURE

2.1 The importance of *Campylobacter* spp.

Campylobacter spp. are Gram-negative, microaerobic, nonsaccharolytic bacteria which are the most common cause of bacterial gastroenteritis worldwide. According to WHO estimates, *Campylobacter* spp. accounted for approximately 27% (96 million cases) of foodborne illness in 2010 (Fig. 1) (3).

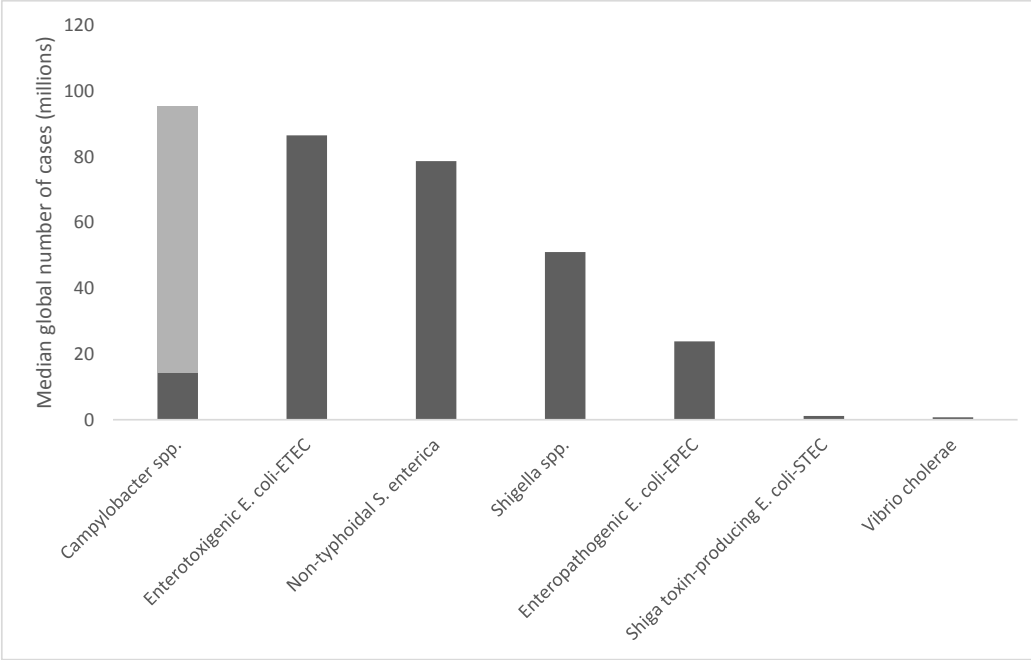

Figure 1. Median global number of cases in millions of bacterial foodborne illnesses in 2010. *Campylobacter* spp. bar represents estimated number of *C. jejuni* (light grey) and *C. coli* (dark grey) cases Source: (3)

In general campylobacteriosis is an acute gastroenteritis disease characterized by abrupt abdominal pain, diarrhoea ranging from watery to bloody, fever, and headache (23). The disease is frequently self-limiting, resolving typically within ten days and with a very low-mortality rate (3, 24). However, in some cases post-infection complications, such as the Guillain-Barré syndrome (GBS), may occur. In poliomyelitis-free countries, GBS is the most common form of acute neuromuscular paralysis (25, 26). In 2010 around 100 000 campylobacteriosis patients (or ~0.1% of the gastroenteritis cases) developed GBS three weeks after *Campylobacter* infection (27-29), implicating that between 28-45% of GBS cases are due to *Campylobacter* infection (3). As a result, among the most common foodborne diarrhoeal disease agents, *Campylobacter* spp. was the leading cause of years lived with disability (YLDs) in 2010 (Fig. 2) (3).

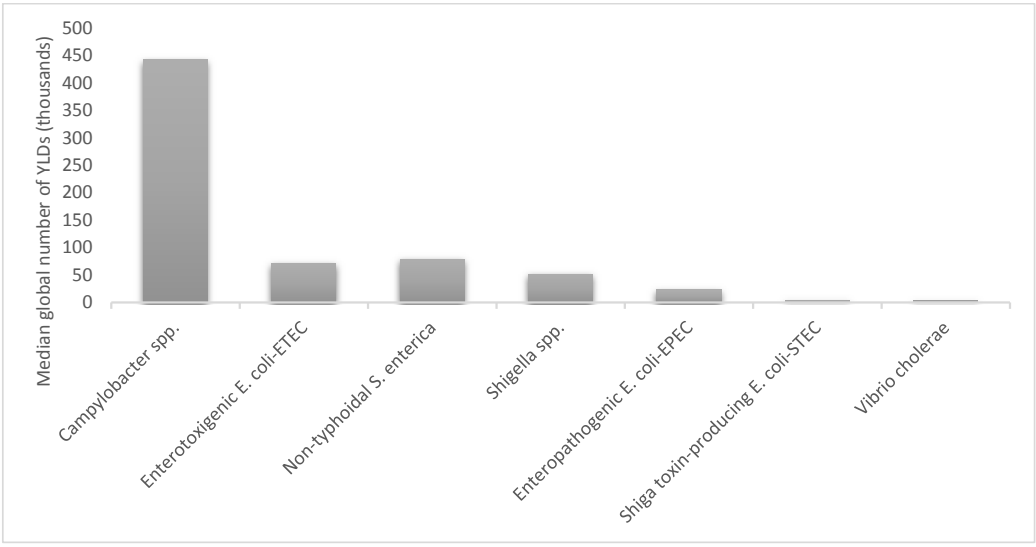

Figure 2. Median global number of YLDs of bacterial foodborne illnesses in 2010. Source: (3)

2.1.1 Campylobacter spp. and GBS

GBS is an atypical autoimmune disease, first described by G. Guillain and J. A. Barré in 1916 (30). In general, GBS is a polyradiculoneuropathy characterized by an acute progressive and symmetrical motor weakness of the extremities with varying degrees of areflexia. Additionally, clear patient-to-patient variation has given rise to GBS subforms such as the Miller Fisher syndrome (MFS) (25, 31). Although GBS was promptly recognized as a post-infectious disease, the aetiological agent remained unknown until 1982 when *C. jejuni* was first isolated from a GBS patient stool (32). Further immunochemical and nuclear magnetic resonance (NMR) studies on *C. jejuni* (33-35) and reproduction of the disease in animal models (36-38) established *C. jejuni*-GBS as the first confirmed case of human autoimmune disease caused by ganglioside mimicry. Gangliosides are glycolipids primarily found in the nervous system and generally anchored in the external leaflet of the lipid bilayer by a ceramide moiety, while extracellularly exposing a highly variable sialylated oligosaccharide moiety (39).

Campylobacter sialylated glycoconjugates have been established as the main structures responsible for the mimicry (11, 38, 40, 41). The vast majority of neuropathogenic *Campylobacter* isolates express variable glycoconjugate structures, resulting in the production of more than one type of antiganglioside antibodies by the patients (42). At the acute stage of GBS, high titres of more than 20 different anti-ganglioside antibodies have been identified. Among those, antibodies against LM1, GM1, GM1b, GM2, GD1a, GalNac-GD1a, GD1b, GD2, GD3, GT1a, and GQ1b gangliosides are the most frequently encountered (8, 36, 41, 43-45). Despite the clear implication of ganglioside mimicry in the pathogenesis of *Campylobacter*-associated GBS/MFS, strains expressing ganglioside-like structures have been isolated from uncomplicated enteritis patients. Thus, a host's genetic

background and immunological status may play an important role in the development of the disease (8, 9, 46).

2.2 *C. jejuni* and *C. coli* population structure and source attribution

The *Campylobacter* genus is currently comprised of 29 species that have been isolated from various species of wild and domesticated birds, mammals, and reptiles. Some *Campylobacter* species are human pathogens (47, 48). Among them, *C. jejuni* is the most frequently isolated species from human patients (~80-97%), followed by *C. coli* (~3-20%) (49-55). *C. jejuni* and *C. coli* are genetically closely related, and seemingly share similar niches, epidemiology, and risk factors (50, 56). The overrepresentation of *C. jejuni* and the apparent overlap between the two species has resulted in *C. coli* being vastly neglected in *Campylobacter* research. Yet, the number of *C. coli* infections surpasses that of other major pathogens (3, 54) (Fig. 1). Recent studies on aetiology (49, 50, 55, 57), epidemiology (53, 58), and genetics (59-62) have shown that *C. jejuni* and *C. coli* are more different than previously thought (63, 64). Additionally, *C. jejuni* and *C. coli* exhibit distinct population structures, as shown by multilocus sequence typing (MLST) data (63-69). MLST is a molecular typing method based on sequence variability of seven unlinked chromosomal loci-encoding genes belonging to the core genome, which are not subject to positive section and that in fact demonstrate selection against amino acid substitution (68, 69). The method assigns an allele number to distinct sequences at each locus. A sequence type (ST) is defined as a unique combination of allelic variants (70). STs that occupy a central position in a population are defined as central genotypes (67, 71). ST groups sharing at least four identical alleles with a central genotype are grouped into a clonal complex (CC). Conversely, a ST that cannot be allocated to a group is considered to be a singleton.

A key aspect of *C. jejuni* and *C. coli* ecology is their ability to colonize multiple hosts (72-76). However, the relationship between host niche and lineage structure in these two species is considerably different (65, 66, 74, 75). *C. jejuni* population is weakly clonal and currently structured into 44 CCs (https://pubmlst.org/campylobacter/ as of May 2017). In general, *C. jejuni* population structure is shaped by niche (i.e. host) segregation. The most evident segregation has been described in wild birds, as only 1% of wild birds STs occur in more than one species (74, 75, 77). Furthermore, *C. jejuni* populations from closely taxonomically related hosts showed low levels of genetic differentiation in MLST genes despite geographic origin (74). Similarly, host-association in agricultural related strains appears to be stronger than phylogeography (78). Some lineages, such as ST-257 and ST-61 CC, exhibited gradual divergence due to host specialization (75, 78, 79). However, the genotype-host association is weaker in farmed animals in comparison to wild birds (75, 78, 79). A large proportion of isolated strains from agricultural sources belong to CCs with a broad host range, thus, these CCs are considered to be host generalist (75, 79-82). ST-21 and ST-45 CCs are two of the most common generalist CCs isolated worldwide (81, 83, 84). Their ubiquity and rapid host jump hamper source attribution efforts and phylogenetic associations (82).

C. coli was originally considered to be less variable than *C. jejuni* according to MLST analysis (59, 69, 85). Early studies found that the majority of *C. coli* (~98%) belonged to two related clonal lineages (59, 85). Nevertheless, these studies had solely focused on MLST diversity in agricultural associated *C. coli*. Typing of *C. coli* from larger and more diverse isolate collections revealed that *C. coli* population is remarkably different from *C. jejuni,* as it is divided into three deep branched clades, containing a total of five distinct lineages (63, 64, 66, 67, 84, 86). Clade 1, encompassing clusters A, B, and C (C1-A, C1-B, C1-C) is primarily associated with agriculture. Meanwhile, clades 2 and 3 are mainly associated with waterfowl and environmental sources (63, 66, 86). Hence, the genetic substructure of *C. coli* is related to ecology rather than host. Furthermore, maintenance of a 3-deeply branched clade population structure and the relatively inexistence of horizontal gene transfer (HGT) between clades suggests three distinct genetic pools (63, 64, 66, 87).

Only two CCs have been defined in *C. coli* clade 1 to date; ST-828 (C1-A) and ST-1150 (C1-B). ST-828 is by far the most abundant CC representing 60-80% of genotyped isolates, while ~2% of the isolates belong to ST-1150 CC (67, 84, 86). Clade 1 shows evidence of recent clonal descent, and is characterised by low synonymous sequence variation at MLST loci and heavy reliance on recombination to generate diversity (63). Recombination in this clade transcends species, as studies have shown a recent increase in asymmetric gene flow between *C. coli* C1-A and C1-B and agricultural *C. jejuni,* possibly enhanced by co-colonization of agricultural hosts (63, 66, 87). Introgression of *C. coli* C1-A and C1-B with the so called agricultural *C. jejuni* population has resulted in whole-allele replacement (63) and occurrence of mosaic alleles (87) in housekeeping genes. Additionally, large proportions of the genetic material throughout the genomes of C1-A (9.7-11.2%) and C1-B (20.4-22.5%) isolates have been shown to be imported from agricultural *C. jejuni* strains (66). This level of introgression has resulted in an average of 9.8 protein coding differences per gene (66). Moreover, if levels of genetic import from agricultural *C. jejuni* remain high, *C. coli* C1-A and C1-B may potentially converge or despeciate (63, 66).

Although the exact impact of *C. coli* clade 1 introgression on agricultural niche adaptation is yet to be studied, the apparent absence of unintrogressed *C. coli* (C-2 and C-3) in farm animals has prompted speculation that import of *C. jejuni* DNA has been crucial for transmission, adaptation, and proliferation in this niche (66). For example, introgression has resulted in an increase in *C. coli* ST-828 CC pan genome size due to the acquisition of new metabolic pathways. Some of these pathways have been shown to provide colonization advantages in *C. jejuni* (66, 88). Yet, *C. coli* ST-828 genome has a smaller number of accessory genes in comparison to *C. jejuni* generalist ST-21 and ST-45 CCs, thus, suggesting more restricted metabolic capabilities (82). This limited metabolic plasticity is apparent by *C. coli* ST-828 CC estimated lower migration rate and zoonotic transmission; one hosts jump every 12 years and lower host switching frequency between mammals and chicken (82). Conversely, *C. jejuni* ST-21 and ST-45 CCs are predicted to jump hosts every 1.6 or 1.8 years, respectively, and show homogeneity in the relative rates of transmission between host species (82).

Although *C. coli* human infections are mainly associated with agricultural sources (53, 64, 89-91), a recent study found clade 3 strains to be overrepresented in clinical isolates (86). This, in addition to previous findings (50, 57), suggests that environment and non-

agricultural food sources (e.g. game animals) may play a significant role in *C. coli* human infections (86). Nevertheless, due to a bias in sampling, little is known about the host, genetic makeup, and potential virulence factors of *C. coli* clade 3.

2.3 *C. jejuni* and *C. coli* glycobiology: from gene to glycan

Glycans and glycoconjugates are the predominant bacterial cell surface structures. As a result of their location at the host-microbe interface, glycosylated structures are involved in shaping bacterial ecology by dictating interactions with microbes, host (92), bacteriophages (93), and the environment in general (94, 95). Although glycan assembly is not template driven, identification and classification of carbohydrate biosynthesis genes offers insights into structures potentially expressed by bacteria.

2.3.1 Identification and annotation of carbohydrate-active enzymes

Although glycan structures do not necessarily correlate with the genetic composition of a glycan biosynthesis locus (7, 96-98), gene content diversity is at the base of glycans structural heterogeneity. Therefore, identification of genes associated to carbohydrate synthesis and glycan modification can theoretically hint at the presence of sugar residues of interest.

There are two general approaches for gene prediction—*ab initio* and homology-based prediction. A*b initio* is an intrinsic method that searches for statistical patterns and DNA motifs within genes and their boundaries. Open reading frames (ORF) are sequences flanked by a translation initiation and termination site. ORF identification and inspection is the most basic *ab initio* method. Since not all ORFs are coding regions, Markov model-based algorithms are used to infer gene description statistically. Conversely, homology-based prediction, is based on the identification of sequences that evolved from a common ancestral sequence. This method assumes a higher level of conservation in coding sequences in comparison to non-coding regions (99). Similarly, functional annotation by sequence homology relies on the premise that during speciation the functionality of corresponding proteins is kept even between species. As a result, a putative function can be assigned to an uncharacterized protein if the activity of a protein with significant sequence similarity is known. Alignment of the query sequence against every sequence in the database is the first step in searching for homology candidates. The heuristic Basic Local Alignment Tool (BLAST) is the most commonly used tool for querying large sequence databases, which can generate either gapped alignments or position-specific iterated (PSI-BLAST) (100). The best BLAST hit (BBH) or the best informative BLAST hit (BIBH) are commonly assumed to be phylogenetically related to the query. However, when comparing distantly related species BBH/BIBH genes might not necessarily be homologous, resulting in misannotation (101). Thus, by predefining a cut-off, coding sequences groups aggregating BBHs sharing high homology and potentially function can be formed. These groups are usually called group of orthologues. Orthology is a subclassification of homology, and it generally refers to genes separated due to speciation (102). Group of orthologues are built using e.g. Markov

Cluster Algorithm (MCL) (103) analysis implemented through different tools such as OrthAgogue (104) and Roary (105).

Although homologues are frequently assumed to encode proteins with similar catalytic properties, functional assignment based solely on homology may be unreliable. Divergent evolution can result in genes encoding proteins with differing specificities (7, 96, 106), or proteins that utilise the same type of intermediate but catalyses different reactions (107), or the genes may have evolved to encode proteins with both mechanism and substrate specificity (108). Furthermore, orthologue cluster delineation and subsequent annotation is further hurdled in multidomain proteins, as homology between single domains might result in the convergence of two protein families. Therefore, to improve annotation protein signature recognition methods, including profile-based, Hidden Markov Model (HMM), and positional specific score matrix can be implemented through e.g. InterProScan (109).

2.3.1.1 Classification of carbohydrate related genes

Oligo and polysaccharides are the most diverse structures in nature, with a calculated diversity of over 10^{12} for hexasaccharides alone (110). The Carbohydrate-Active enZymes (CAZymes) are the enzymes involved in the biosynthesis and breakdown of glycans. Currently, the CAZy database (CAZyDB) is the most comprehensive catalogue of CAZYmes with almost 400 000 unique protein sequences (http://www.cazy.org/) (12) classified into five functional classes; glycoside hydrolases (GHS), glycosyltransferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and auxiliary activity enzymes (AA) (12). The GTs functional class encompasses enzymes catalysing glycosidic bond formation using activated sugars as donors and sugars, lipids, protein, nucleic acids, antibiotics, or other small molecules as acceptors. Despite the considerable variety of donors, staggering number of acceptors, and the cosmic diversity of products, GTs most likely underwent divergent evolution from few common ancestors, as reflected by the two types of structural folds displayed by nucleotide sugar-dependent GTs: GT-A and GT-B (111-113). As a result, functional annotation of GTs is a highly challenging task. Currently, GTs are classified into 103 families based on amino acid sequence similarity (12, 111, 114). This type of classification permits systematic and logical grouping of GTs into families, without requiring demanding functional studies, and provides insights into the families' divergent evolution. Since amino acid sequence similarity is strongly correlated with protein folding, this classification allows grouping enzymes with similar three dimensional structures and molecular mechanisms. However, this classification has been shown to poorly correlate with enzymes specificities, as polyspecificity within a family is common (12, 114). Moreover, several studies have shown that substitution of a single or very few amino acids can modify the activity of a particular GT (7, 10). Consequently, functional predictions and annotations beyond family classification should be discouraged if the biochemical properties of the GT have not been established (98) and/or if the enzyme in question is not highly similar to the members of family extensively characterized as monospecific. To identify, classify, and annotate GTs, blasting the genome against the

annotated full-length CAZyDB is a common practice. However, this method requires manual curation of the data, and CAZy family assignment is only transferred to the most significant hit (115). Thus, automatic annotation methods based on HMM (115, 116) or on peptide pattern recognition (PPR) (117) have been developed. Yet, differences in the sensitivity and precision of the various methods have been observed (115, 117). In *C. jejuni* and *C. coli,* 15 GT families are represented among the 138 strains listed in the CAZy database as of April 2017. However, biochemical characterization has only been carried out for 12 different *C. jejuni* glycosyltransferases (7, 10, 96, 98, 106, 118-125).

2.3.2 *C. jejuni* and *C. coli* carbohydrate gene clusters

Glycans have a fundamental role in the biology and pathogenesis of *C. jejuni* and *C. coli* (4, 5), and a significant proportion of the genome (~10%) is estimated to encode GTs and other enzymes involved in the synthesis and modification of sugar nucleotides and glycosylated structures (6). In *Campylobacter* genomes, these carbohydrate-related genes are mainly clustered in four distinct loci;

(i) O- linked protein glycosylation pathway

Protein *O*-glycosylation, a widespread pathway in bacteria, is the glycosylation of serine or threonine residues at the hydroxyl group commonly carried out sequentially on the surface exposed residues by GTs. In *C. jejuni* and *C. coli O*-glycosylation of flagella proteins FlaA and FlaB heavily contributes to the molecular mass of flagellin (~10%) and is essential for flagella assembly (126, 127). Nevertheless, the nature of flagellar glycans is strain specific and determined by the genetic content of the hypervariable *O*-glycosylation locus, which is located contiguous to the flagellin structural genes (6). For example, *C. jejuni* 81-176 has the smallest *O*-glycosylation locus reported so far with only 21 genes, some of which encode enzymes that synthesize 5,7-diacetamido-3,5,7,9-tetradeoxy-L-glycero-α-L-manno-nunolusonic acid (pseudaminic acid, Pse5Ac7Ac) and related derivatives (127-129). *C. jejuni* 81-176 FlaA and FlaB are predominately glycosylated with Pse5Ac7Ac and 5-acetamido-7-acetamidino-3,5,7,9-tetradeoxy-L-glycero-α-L-manno-nunolusonic acid (Pse5Ac7Am), and minor amounts of *N*-acetylglutamine (Pse5Ac7AcO-GlcAc) and *O*-acetyl (Pse5Ac7AcOAc) (127, 128, 130, 131). In contrast, *C. jejuni* NCTC 11168 *O*-glycosylation locus encompasses almost 50 genes, including a gene cluster for 5,7-diacetamido-D- glycero-β-D-galacto-nunolusonic acid (legionaminic acid, Leg5Ac7Ac) synthesis and related derivatives (6). Consistently, *C. jejuni* NCTC 11168 flagellin proteins are glycosylated with Pse5Ac7Ac, Leg5Ac7Ac, and derivatives, with highly heterogeneous glycosylation patterns (Fig. 3) (132-135). Similarly, *C. coli* VC167 glycosylates its flagellin with Pse5Ac7Ac, Pse5Am7AcO-deoxypentose, Pse5Ac7AcO-deoxypentose, acetamidino (Leg5Am7Ac), and *N*-methylacetimidoyl Leg5AmNMe7Ac of Leg (136, 137). Hence, *O*-glycosylation is genetically and phenotypically highly heterogeneous. This diversity has significant impact on the colonization and virulence of campylobacters (135).

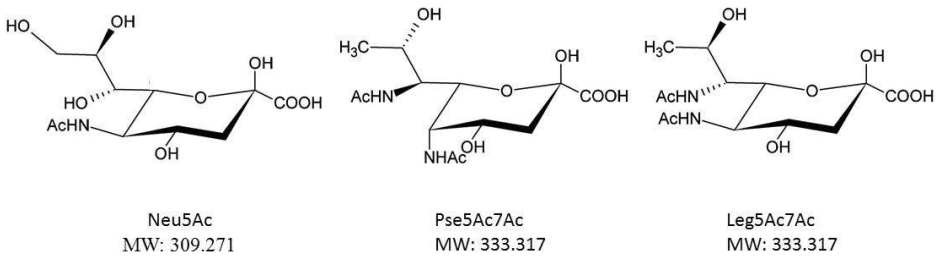

Figure 3. Structures of sialic acid and sialic acid-like sugars. Modified from (138)

(ii) N-linked protein glycosylation systems

Protein *N*-linked glycosylation is the addition of a glycan to a nitrogen atom of an asparagine residue. In *C. jejuni* and *C. coli,* the heptasaccharide *N*-glycan is a remarkably conserved GalNAc-α1,4-GalNAc-α1,4-[Glcβ1,3-]GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac-β1,N-Asn, where Bac is bacillosamine, 2,4-diacetamido-2,4,6-trideoxyglucopyranose (139-141). Similar to eukaryotes, this glycan is linked to proteins through the asparagine in the Asn-Xaa-Ser/Thr motif (139). However, the *en bloc N*-glycosylation occurs on the surface-exposed parts of a folded protein instead of on the unfolded protein. All the highly conserved enzymes required for the synthesis and transfer of the *N*-glycan are found within the 17 kb protein glycosylation locus (*pgl*) located adjacent to the LOS biosynthesis locus (6, 141-143). The *N*-glycosylation pathway has been elucidated (144) and several of the proteins involved in the pathway have been fully characterized (121-124). Impairment of the *N*-linked glycosylation glycosylation pathway leads to pleiotropic phenotypes in *C. jejuni*, resulting in a general decrease in bacterial fitness (145-147).

(iii) Capsule

In the early years of molecular research, *C. jejuni* and *C. coli* were thought to synthesize low and high molecular weight glycolipids inferred to be lipopolysaccharides (LPS). These structures were assumed to be the heat stable antigens that differentiated strains in the Penner serotyping schema, despite evidence otherwise (148). This paradigm was generally undisputed until the publication of the first *C. jejuni* genome and the identification of genes putatively involved in capsule (CPS) transport and assembly (*kps* genes) (6). Mutational studies of these genes demonstrated that indeed the high molecular weight glycolipid was a CPS (149). Additionally, Penner serotype was shown to rely on CPS structural variability for strain classification. Bacterial CPS is a polysaccharide chain made of variable length containing repeating units of one or more sugar residues, and which is loosely linked to the cell surface by a lipid anchor. Based on the biosynthesis mechanism, assembly, genetic regulation, and sequence similarity, CPS can be broadly divided into four groups (I-IV). In CPS group II and III gene loci, the highly conserved transport and assembly (*kps* genes) genes flank a hypervariable biosynthetic region. In general the gene order is *kpsFEDUCS…TM* for group II and *kpsDMTE...CS* for group III cluster. *C. jejuni* and *C. coli* have a II/III hybrid capsule type as the gene content and order is *kpsMTEDF….CS.*

Although genes outside the CPS locus may contribute to CPS structural variability, the genetic diversity within the hypervarible CPS region forms the basis of CPS heterogeneity (150, 151). Comparative genomics have identified 12 CPS gene content classes in *C. jejuni* (A-L) and eight in *C. coli* (A-H) (17, 150).

(iv) Lipooligosaccharides

LPS are the most abundant glycoconjugates found in Gram-negative bacteria outer leaflet of the outer membrane. These phosphorylated endotoxin glycolipids help to modulate host immune response, maintain structural integrity, and cope with environmental stresses (152, 153). LPS is typically composed of a hydrophobic glucosamine-based lipid A anchor substituted with a relatively conserved core polysaccharide, and a highly variable repeating O-polysaccharide chain (154). A rough version of LPS devoid of the O-polysaccharide chain known as lipooligosaccharides (LOS) is found in *C. jejuni* and *C. coli,* as well as in some mucosal pathogens such as *Neisseria* and *Haemophilus* (155). In humans, free LPS and LOS is recognized and bound by the LPS-binding protein (LBP), which delivers the glycolipid to cluster of differentiation 14 (CD14). Then, CD14 presents the LPS/LOS to Toll-like receptor 4-myeloid differentiation factor 2 (TLR4-MD2). An intracellular signalling cascade is initiated upon binding of LPS/LOS to TLR4-MD2, triggering the innate immune response (156-158). Although only the lipid moiety of LPS/LOS is required for initiating the innate immune response, structural differences in the glycan fraction have been shown to have a modulatory effect (159-164). The genetics and structure of *C. jejuni* and *C. coli* LOS are extensively presented in the following chapters.

2.4 Genetic basis of *C. jejuni* and *C. coli* LOS

With the publication of the first *C. jejuni* genome (6) and the advent of whole genome sequencing (WGS), genotypic-phenotypic studies have been conducted to understand LOS structural heterogeneity and to link genetic backgrounds with disease outcomes (7, 46, 97, 165-167). In *C. jejuni* and *C. coli,* LOS biosynthesis genes are clustered within the LOS biosynthesis locus (168, 169). The LOS locus is a hypervariable genetic region in *Campylobacter* spp. (170-173) flanked by the conserved LOS biosynthesis genes *waaC* and *waaF* (7). Various studies have explored the remarkable diversity and organization of gene complement of this locus in *C. jejuni,* which arises from extensive recombination and horizontal gene transfer (7, 17, 174-176). Comparative analysis of *C. jejuni* LOS loci has identified 40 LOS orthologues arranged in 22 gene content classes named A-W (excluding N) (7, 17, 174, 175). These gene content classes, known as LOS locus classes, have a mosaic nature as they have arisen by the insertion and/or deletion of genes or gene cassettes (7, 175). Similar to other virulence genes (177), a strong link between *C. jejuni* LOS locus classes and CC/STs has been observed (167, 178-181). Thus, LOS class distribution is associated with *C. jejuni* population structure (i.e. same lineages harbours same LOS classes).

In contrast, the genetic makeup of the *C. coli* LOS biosynthesis locus was largely unknown until recently (17). The *C. coli* LOS locus was found to be more genetically diverse, as 51 distinct orthologues were identified. Nevertheless, these orthologues were arranged in eight

LOS locus classes named I to VIII (17). As in *C. jejuni*, *C. coli* LOS classes appear to be non-randomly distributed in nature, as a strong association between LOS locus class and host was observed (17). Additionally, introgression of *C. coli* with *C. jejuni* was identified as an important genetic mechanisms for generating LOS diversity, and potentially influencing virulence (17).

2.4.1 Genetic mechanisms generating *C. jejuni* and *C. coli* LOS variability

In addition to gene content diversity, *C. jejuni* utilizes at least four other mechanisms to generate LOS structural heterogeneity (7, 96, 169, 175). Mechanisms and their respective examples are shown below to illustrate how minor genetic differences can have large repercussions on LOS structures.

(i) Phase variation due to homopolymeric tracts

Phase variation of surface antigenic structures is an important adaptive strategy in which reversible hypermutation or hypervariable methylation generates population diversity, which enhances bacterial survival (182, 183). Simple sequence repeat (SSR) tracts are a phase variation mechanism involving iterating units of one to several nucleotides. Mononucleotide tandems, known as homopolymeric tracts (HTs), are the simplest SSR commonly located within the reading frame or in the promoter region of a gene. HTs are highly prone to insertions or deletions during processes requiring DNA synthesis due to a slipped-strand mispairing mechanism. Consequently, the length of HTs is altered, producing reversible translational frameshift mutations in the coding sequence of a gene (182-184). Therefore, within a clonal population a mixture of translationally in- and out-of frame gene variants may be present (7). HTs between 5-13 nucleotides are commonly found in *C. jejuni* carbohydrate biosynthesis genes (6). While poly(A) and poly(T) HTs are overrepresented in the coding genes of several prokaryotes, poly(C) and poly(G) are the most common HTs in *C. jejuni* (6, 185), while a sole instance of phase variable poly(A) HT within the LOS loci has been reported (186). LOS structural heterogeneity due to phase variation linked to HTs was first demonstrated in *C. jejuni* NCTC 11168 (169). An intragenic poly(G) HT of either eight or nine bases was identified in the gene *wlaN*. The 8-poly(G) variant encoded a function β-1,3-galactosyltransferase, while the 9-poly(G) expressed an inactive truncated version. Since WlaN transfers the terminal β-1,3-Gal in the GM1a LOS mimic, inactivation of this protein results in the expression of a GM2-like LOS. *C. coli* has been suggested to have HTs-linked phase variability potentially similar to *C. jejuni* (17). However, LOS structural heterogeneity due to HTs has not been examined in *C. coli*. In fact, the impact of HTs on the expression of surface structures in *C. coli* is limited to a single study in which a phase variable poly(T) HT in the *flhA* gene was responsible for immotile cells (187).

(ii) Deletion/insertion of a single base without phase variation

C. jejuni OH4384 and OH4382 carry identical 11.5 kb LOS locus class A, with the exception of a single base deletion at position 71 of *orf5*. This deletion causes a frameshift mutation, effectively inactivating the 1,4-*N*-acetylgalactosaminyltransferase encoded by *orf5* (7, 96). As a result, *C. jejuni* OH4382 expresses a GD3-like LOS instead of the GT1a mimic synthesized by *C. jejuni* OH4384.

(iii) Loss of enzymatic activity by a single amino acid mutation

The LOS locus class C *orf5/10* of *C. jejuni* NCTC 11168 encodes a β-1,4-N-acetylgalactosaminyltransferase/CMP-NeuAc synthase fusion which is essential for the synthesis of GM2 and GM1a mimics (7, 188). *C. jejuni* ATCC 43430 (LOS locus class C) has a single amino acid substitution (C92Y) in this gene, which abolishes the β-1,4-N-acetylgalactosaminyltransferase activity. Consequently, *C. jejuni* ATCC 43430 only expresses a GM3-like LOS structure (7).

(iv) *Multiple mutations leading to differences in acceptor specificity*

Variants of the β-1,4-N-acetylgalactosaminyltransferase CgtA found in LOS locus classes A, B, and C share a 34% overall amino acid sequence identity (7). As a consequence of this significant sequence divergence, a marked difference in acceptor preference has been observed between the variants (7, 96). For example, the CgtA from *C. jejuni* ATCC 43438 (LOS class A) is only active on non-sialylated acceptors. Conversely, OH4384 (LOS class A) is only active on monosialylated acceptors, and ATCC 43456 (LOS class B), and NCTC 11168 (LOS class C) are active on mono- and disialylated acceptors (7, 96).

2.5 *Campylobacter jejuni* and *C. coli* lipooligosaccharides composition and structure

Research efforts have been directed at elucidating the chemical composition and structure of *C. jejuni* LOS, as several ganglioside mimics have been found in this species. Thus, information concerning LOS chemical structure and composition is largely limited to *C. jejuni*. This is despite the fact that *C. coli* has been isolated from GBS patients, and its role in post-infectious diseases is debated (14-16). Hence, in the following sections the characteristics *C. jejuni* LOS will be presented in detail, while *C. coli* data will be limited to the few isolates for which the LOS chemical composition is known (18, 19) and the single chemical structure that has been resolved (189).

2.5.1 Lipid A

Lipid A is the most structurally conserved moiety in both LPS and LOS, and it is essential for both viability and maintaining the membrane barrier (154, 155). Structural variations are limited to disaccharide backbone composition (D-glucosamine (GlcN) and/or 2,3-diamino-2,3-dideoxy-D-glucose (GlcN3N)), degree of phosphorylation and phosphate substituents, and the length, number, and linkage of the fatty acid acyl chains (92, 143, 154, 155, 190-192).

Early structural studies in *C. jejuni* LOS revealed a heterogeneous lipid A backbone composed of three β(1→6)-linked hexosamine disaccharides (193), considerably distinct from *C. fetus* lipid A (194). Although differences in the relative proportion of the various disaccharide species have been observed between and within *C. jejuni* strains (143), GlcN3N-GlcN is generally the most abundant species followed by GlcN3N-GlcN3N and GlcN-GlcN (143, 164, 193, 195). Despite saccharide compositional differences, all three hexosamine disaccharide backbones are phosphorylated and acylated in a similar fashion,

giving rise to a biphosphorylated hexacylated lipid A. Disaccharides are phosphorylated at the hydroxyl group at C-4' of the non-reducing hexosamine and at the glycosidic hydroxyl group at C-1 of the reducing hexosamine (143, 164, 193). Additionally, phosphate groups can be further modified with a phosphoethanolamine (PEtn) residue by the action of the promiscuous pEtN transferase encoded by *cj0256* (*eptC*) (196), yielding pyrophosphoethanolamine (PPEtn) (143, 164, 193). This modification contributes to an increase in resistance to polymixin B and avian and human cationic antimicrobial peptides (CAMPs), and modulates the innate immune response (164, 197).

Simultaneously, the dissacharide backbone is asymmetrically substituted with six fatty acid chains. At positions C-2 and C-3 the disaccharide backbone is substituted with (*R*)-3-hydroxytetradecanoic acid [14:0(3-OH)], while (*R*)-3-(hexadecanoyloxy)tetradecanoic acid [3-(16:0-*O*)-14:0] or (*R*)-3-(tetradecanoyloxy)tetradecanoic acid [3-(14:0-*O*)-14:0] substitute at C-2', and 3-(16:0-*O*)-14:0 at C-3' (193, 195). Hence, fatty acid distribution in *C. jejuni* lipid A is in general isostructural to *E. coli* lipid A, but with a partial replacement of tetradecanoic acid with hexadecanoic acid (154, 198). Actually, *C. jejuni* lipid A structurally (198) resembles the general enterobacterial lipid A aside from the existence of GlcN3N in the backbone. The presence of GlcN3N in the disaccharide backbone results in an increased proportion of amide linked acyl chains. Initially, this structural difference was regarded as inconsequential to *C. jejuni* LOS antigenic activity (20, 193, 199, 200). However, upon the characterization of the enzymes essential for the synthesis of GlcN3N substituted lipid A, GnnA and GnnB, in *Acidithiobacillus ferroxidans* (201) and the identification of these protein homologues in *C. jejuni* 11168 (Cj0504c and Cj0505c, 40% and 55% amino acid similarity, respectively), the true impact of GlcN3N on *C. jejuni* LOS antigenicity was recognised (202). Deletion of either *C. jejuni* 11168 *gnnA* or *gnnB* (Cj0504c and Cj0505c, 40% and 55% amino acid similarity to *A. ferroxidans*, respectively) resulted in the addition of an ester-linked acyl chain instead of an amide-linked acyl chain (202). Substitution of ester-linked with amide-linked acyl chains causes a decrease in the antigenic activity of *C. jejuni* LOS on human TLR4-MD2 and tumour necrosis factor alpha (TNFα) expression (164, 202). Thus, variability in the proportion of amide-linked acyl chains has direct impact on the modulation of the innate immune response (164, 202). In contrast to *C. jejuni*, *C. coli* lipid A composition has only being superficially analysed by Naess and Hostad (18), as *C. coli* LOS chemical studies have been solely focused on determining the LOS core composition (18-22). Thus, the nature of the disaccharide backbone, and the degree of acylation and phosphorylation in *C. coli* lipid A has been long assumed to be identical to *C. jejuni*, even though lipid A structural differences have been observed within the *Campylobacter* genus (194).

2.5.2 LOS inner core

Based on the composition of the polysaccharide core, this moiety can be subdivided into inner core and outer core (154). The inner core is the region proximal to the lipid A, and it is distinguished from the outer core by the presence of L-glycero-D-manno-heptopyranose (L,D-Hep), D-glycero-D-manno-heptopyranose (D,D-Hep), and most importantly, 3-

deoxy-D-manno-octulosonic acid (KDO). Regardless of the bacterial species, all LPS and LOS studied thus far contain at least one KDO residue which links the lipid A hexosamine disaccharide backbone at position C-6 with the inner core (154, 155). Similar to lipid A, the backbone structure of the LPS inner core is largely conserved within closely related species such as within the *Enterobacteriaceae* family. In *E. coli,* five core types (R1-R4 and K12) have been described for over 100 different serotypes (203). The tetrasaccharide Glc-α(1,3)-L,D-Hep-α(1,3)-L,D-Hep-α(1,5)-KDO is the typical enterobacterial inner core. Nevertheless, heterogeneity in the inner core may arise from nonstoichiometric substitutions with residues such as rhamnose (Rha), galactose (Gal), GlcN, N-acetylglucosamine (GlcNAc), KDO, phosphate and phosphorylethanolamine (PEtN). In contrast to LPS, LOS inner core is less conserved among different species (155). All *C. jejuni* studied to date have a LOS basic inner core composed of L,D-Hep-α(1,3)-L,D-Hep-α(1,5)-KDO, where the L,D-Hep adjacent to KDO is substituted with a D-glucose in β(1,4) linkage (8, 33, 34, 143, 195, 204-209). Although the typical *Enterobacteriaceae* tetrasaccharide inner core has been found in some *C. jejuni* strains, the vast majority of the strains have a Gal residue β(1,3) linked to the second L,D-Hep (8, 33, 34, 143, 204-209) instead of a Glc residue (8, 204). Additionally, the inner core region can be further derivatized by the action of EptC. As in the lipid A, EptC transfers a pEtN residue to the innermost Hep (152, 197). The influence that this inner core modification has on virulence and host-pathogen interaction is unknown due to the pleiotropy of EptC (197). The single available *C. coli* LOS inner core resembles *C. jejuni* basic inner core, with the exception of a further substitution of β(1,4) Glc with a β(1,3) Hep and the absence of phosphate substituents in the inner core (21).

2.5.3 LOS outer core

The outer core is the most variable region in *C. jejuni* LOS, and with greater interstrain variability than that observed in *E. coli* (210). Typically, LPS or LOS outer core regions are composed of D-glucose (Glc), D-galactose (Gal), GlcNAc, and *N*-acetyl-D-galactosamine (GalNAc) (154, 155). However, chemical and structural studies of the *C. jejuni* outer core revealed the presence of two uncommon features—*N*-acetylneuraminic acid (Neu5Ac, sialic acid) (8, 33, 34, 143, 204, 206, 207) and 3-amino-3,6-dideoxy-D-glucose (Qui3N), commonly found as a 3-acetamido-3,6-dideoxy-D-glucose (Qui3NAc) residue (8, 143, 205, 208, 209). Based on the available chemical studies on *C. jejuni* LOS, the presence of Neu5Ac and Qui3NAc appear to be mutually exclusive (8, 33, 34, 143, 204-209). Neu5Ac is the most abundant nonulosonic acid found widespread on eukaryotic surface glycans, but relatively uncommon in prokaryotic organisms (211, 212). Conversely, Qui3NAc is an unusual dideoxyamino sugar which thus far has been exclusively reported in a limited number of bacterial surface structures. Qui3N has been found in all *C. coli* strains analysed to date, suggesting that this residue is characteristic of *C. coli* LOS (18-22). In contrast to *C. jejuni*, Qui3N is found in *C. coli* in an *N*-acylated form (Qui3NAcyl), with either (*R*)-3-hydroxybutanoyl or 3-hydroxy-2,3-dimethyl-5-oxoprolyl as *N*-acyl groups (21). Interestingly, Qui3NAcyl residues were found to cap *C. coli* LOS (21). Thus, it is tempting

to speculate that Qui3NAcyl may play an important role in many biological interactions due to its ubiquity and position in *C. coli* LOS structure.

2.6 Biosynthesis of Sialyl-glycoconjugates

Nonulosonic acids are a highly diverse structural class of nine-carbon α-keto acids (213). Nevertheless, the presence of nonulosonic acids in glycoconjugate has been shown to be key in the pathogenesis of various bacterial species (214). There are two subclasses of nonulosonic acids: the 5,7-diamino-3,5,7,9-tetradeoxy-non-2-ulosonic acids, which includes legionaminic and pseudaminic acids; and the 5-amino-2-keto-3-deoxy-5-acetamido-D-glycero-D-galacto-nonulosonic, known as sialic acid (acetylneuraminic acid, Neu5Ac). First discovered in 1936 in the bovine submaxillary mucin, hence the Greek name 'sialon', sialic acid is the most widely distributed nonulosonic subclass (215, 216). This subclass is a family of over 50 naturally occurring sugars derived from Neu5Ac and 2-keto-3 deoxy-D-D-glycero-D-galcto-2-nonulosonic acid (KDN), the former being the most common sialic acid in nature (212, 217). Sialic acids are often found capping the outer ends of glycans on all vertebrate cell surfaces and in the cell surface of some bacteria (218). Frequently, these bacteria are pathogenic and sialylation of surface structures allows them to circumvent the immune system (214, 219, 220). Bacteria utilize four different mechanisms for cell surface sialylation: *de novo* synthesis, *trans*-sialidase, and donor or precursor scavenging (219, 221).

For LOS sialylation, *C. jejuni* requires genes encoding enzymes involved in *de novo* synthesis of activated Neu5Ac and transfer (see below). Based on their genetic content, only *C. jejuni* strains carrying an A, B, C, M, R, or V LOS locus class can potentially exhibit a sialylated LOS structure (7, 17, 174, 175). In fact, the LOS from representative strains from all these classes, except V, have been shown to be sialylated (7, 9, 10, 96, 222). LOS locus classes, A, B, and, C have been of particular interest due to their association with ganglioside biosynthesis and GBS (9, 46, 165, 223, 224). In contrast, none of the *C. coli* LOS classes reported to date carry all of the genes associated with the expression of a sialylated LOS structure (17). Therefore, the following sections will be limited to the synthesis and transfer of Neu5Ac in *C. jejuni.*

2.6.1 Sialic acid biosynthesis genes

Sialic acids are peculiar sugars synthesized through the condensation of a neutral six carbon molecule with a three carbon pyruvate, and activated with a high energy nucleotide (212). Neu5Ac biosynthesis genes (*neu*) were first identified in region 2 of *E. coli* K1 CPS locus (225). In bacteria, *de novo* Neu5Ac biosynthesis begins with the epimerization of UDP-GlcNac to ManNAc (*N*-acetylmannosamine) by a UDP-*N*- GlcNac 2-epimerase (NeuC) (226). Then, ManNAc is condensed with phosphoenolpyruvate (PEP) to form Neu5Ac by an *N*-acetylneuraminic acid synthetase (NeuB) (227). Finally, free Neu5Ac is activated with cytidine triphosphate to form cytidine 5'-monophosphate (CMP)-Neu5Ac by the action of a CMP-*N*-acetylneuraminic acid synthetase (NeuA) (228). Similar biosynthetic steps have been observed in the synthesis of other nonulosonic acid sugars such as legionaminic and pseudaminic acid, and KDN, and reactions are catalysed by homologous enzymes (131, 138, 211). In the nonulosonic acid biosynthesis pathway the enzyme involved in the

condensation of a 6-carbon molecule with PEP (NeuB in Neu5Ac synthesis) is the most conserved (203). Additionally, partial homology to nonulosonic biosynthesis enzymes has been observed in the KDO biosynthesis pathway. In fact, the structure, biosynthesis (by condensation of D-arabinose with PEP), and activation (CMP-nucleotide) of this bacterial eight carbon α-keto acid is very similar to nonulosonic acids (212, 221, 229).

In *C. jejuni* three NeuB homologues have been identified: Cj1141, Cj1327, and Cj1317 (6, 230). Although all NeuB homologues were shown to restore *E. coli* EV24 sensitivity to K1 bacteriophage, only Cj1141 was shown to synthesize Neu5Ac *de novo* at significant levels (230). Additionally, mutational studies revealed the loss of a GM1-like LOS structure upon deletion of *cj1141,* while *cj1327* and *cj1317* deletions had no effect on LOS structure (230). Consequently, the protein encoded by *cj1141* was confirmed to be a Neu5Ac synthetase (*neuB*) involved in sialylated LOS biosynthesis (230). Later studies identified Cj1327 and Cj1317 as legionaminic (NeuB2, PtmC, LegI) and pseudaminic acid synthetases (NeuB3, PseI), respectively (131, 138).

Downstream from *neuB,* a putative UDP-*N*- GlcNac 2-epimerase (*cj1142*) and a putative CMP-Neu5Ac synthase (*cj1143*) were identified (6, 230). Similar to *neuB,* deletion of *cj1142* resulted in an asialylated LOS structure (231). However, deletion of *cj1143* did not result in the loss of Neu5Ac from the LOS core, possibly due to the activity of NeuA homologues from the legionaminic and pseudaminic acid biosynthesis pathway (231).

2.6.2 Sialyltransferases

Sialyltransferases are key enzymes in the synthesis of sialylated glycoconjugates. Sialyltransferases are inverting GTs which catalyse the formation of an α-sialoside from, typically, the activated donor CMP-β-Neu5Ac (232-237). Characterized sialyltransferases have been shown to catalyse the formation of α2,3-, α2,6-, or α2,3/8-sialyl linkages, while polysialyltransferases synthesize α2,8-, α2,9- or alternating α2,8/α2,9-sialyl linkages (12, 217). Terminal Gal, GalNAc, and Neu5Ac, are the most commonly recognized acceptors by sialyltransferases. In contrast to mammalian sialyltransferases, which are commonly monofunctional and have strict substrate specificity, many bacterial sialyltransferases exhibit sialidase and *trans*-sialidase activities, as well as substrate promiscuity (120, 217, 238, 239). Following the sequence-based classification available at the CAZy server (www.cazy.org), sialyltransferases have been categorized into six glycosyltransferases families (12). Eukaryotes and virus sialyltransferases are restricted to GT-29 family, while prokaryotic sialyltransferases are found in the GT38 (polysialyltransferases), GT-42, GT-52, GT-80, and GT-97 families (217, 238, 240).

So far, three sialyltransferases have been identified in *C. jejuni*: CstI, CstII, and CstIII. All three enzymes belong to the GT42 family, which is currently comprised of enzymes with sialyltransferase, sialidase, and *trans*-sialidase activities.

The CstII from *C. jejuni* was the first chemically and structurally characterized GT42 (96, 119). In solution, CstII is a tetramer in which each monomer is organized into two closely associated domains: the main central domain is a single mixed α/β Rossmann fold containing the nucleotide-binding domain, while the secondary domain is a long coil with

two helices that form a lid-like structure that folds over the active site (118, 119). Since the DxD motif is absent, CstII has a GT-A variant 1 type structure (119). CstII is either a mono or bifunctional α2,3/8-sialyltransferase with α2,8-sialidase and α2,8-*trans*-sialidase activities, and broad acceptor flexibility (120, 236, 241, 242). The bifunctional CstII variant is approximately 93–97% identical to the monofunctional variant (7), and a single amino acid substitution (N51T) has been shown to abolish α2,8-ST activity (7). The bifunctional CstII variant demonstrates good α2,3-sialyltransferase activity on Lac and LacNAc acceptors, and excellent α2,8-activity on α2,3-sialosides and α2,6-sialosides. Additionally, CstII can utilize CMP-Neu5Gc and CMP-KDN as donors, albeit at a lower transfer rate (120, 232).

Similar to CstII, CstI is a tetramer with GT-A variant 1 type monomers. However, each monomer carries an additional C-terminal domain of unknown function (125). Amino acid residues involved in substrate binding and catalysis are conserved between CstI and CstII, despite their overall low sequence identity (42%) (119, 125). CstI has been shown to have good sialyltransferase activity on Gal and Lac type II, and a broad acceptor tolerance (236, 243-246), however the biological role of CstI in *C. jejuni* remains unknown.

3 AIMS OF THIS STUDY

The main aim of this study was to investigate the diversity of *C. coli* LOS biosynthesis locus with particular interest in identifying genetic features associated with the synthesis and transfer of nonulosonic acid. An additional aim was to investigate the presence of nonulosonic acid in *C. coli* LOS.

The specific aims were:

1. To investigate *C. coli* LOS locus diversity. Studies II and III.
2. To examine the impact of horizontal gene transfer in shaping *C. coli* LOS locus. Studies I-III.
3. To identify and assess the distribution of the molecular machinery associated to the synthesis of sialylated LOS. Studies II and III.
4. To examine the role of GT-42 encoding genes in *C. coli* LOS biosynthesis. Study I and IV.

4 MATERIALS AND METHODS

4.1 Bacterial isolates and DNA extraction (I, II, and IV)

A total of 144 *C. coli* isolates (90 swine, 34 humans, 18 poultry, and 2 wild birds) from previous studies (56, 247-252) and from the Campynet (CNET) collection (hosted by DSMZ GmbH [https://www.dsmz.de/]) were selected based on genotype and origin to encompass the greatest possible diversity. Isolates were retrieved from -70 ºC frozen stocks and routinely cultured on Nutrient broth Nº2 (CM0067, Oxoid) with 1.5% (w/v) of agar supplemented with 5% (v/v) defibrinated horse blood (NA plates), unless otherwise specified. All strains were incubated at 37 ºC overnight under microaerobic conditions. DNA was isolated using the Wizard Genomic DNA Purification Kit (Promega, Germany) or using the PureLink Genomic DNA kit (Life technologies, USA). Species was confirmed using species-specific PCR (253).

4.2 Whole genome sequencing and assembly (I-III)

In study I, *C. coli* 76339 genome was sequenced using 454 Titanium (LGC Genomics GmbH, Germany) with a coverage of > 30×. The genome was assembled into a scaffold using a combination of pair-end and 8 kb mate-pair libraries with MIRA 3.2. (254), SSAKE (255), and Staden (256). Scaffolds were verified by PCR and Sanger sequencing.

In study II, either HiSeq or MiSeq were used to sequence the draft genome of a total of 35 isolates. HiSeq library preparation, enrichment, sequencing, and sequence analyses were performed by the Institute for Molecular Medicine Finland (FIMM Technology Center, Finland). MiSeq sequencing was performed by the Institute of Life Science, Swansea University (Swansea, United Kingdom). Reads were filtered and assembled with SPAdes Assembler v. 3.3.0 (257).

In study III, raw sequencing data of a selected group of *C. coli* strains was retrieved from ENA with getSeqENA (https://github.com/B-UMMI/getSeqENA). Using the INNUca pipeline (https://github.com/INNUENDOCON/INNUca), paired-end raw reads were assembled if the raw data fulfilled the expected minimum 15x coverage. Quality analysed reads with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), were cleaned with Trimmomatic (258). Then, *de novo* draft genome assembly was performed with SPAdes 3.11 (257) and the assembly depth of coverage was checked (min 30x). Following draft genome improvement with Pilon (259), species was confirmed and MLST was predicted with *mlst* software (https://github.com/tseemann/mlst).

4.3 Genome annotation (I-III)

All genomes in study I and II were primarily annotated with Rapid Annotation Using Subsystems Technology (RAST) (260). For *C. coli* 76339 genome (study I), Artemis (261) was used to manually curate and re-annotated genes of interest. Gene homology was identified using NCBI's BLAST suite of programs, using UniProtKB/Swiss-Prot (262) as a reference database. Protein conserved functional domains were identified with InterProScan (263). The whole-genome sequences of all *C. coli,* with the exception of 76339, are publicly available on the RAST server (http://rast.nmpdr.org) with a guest account (login and password 'guest') under accession numbers 195.91, 195.96 to 195.119, 195.124 to 195.126, 195.128 to 195.130, 195.133, 195.134, and 6666666.94320. The genome of *C. coli* 76339 was deposited in EMBL with the accession number HG326877.

In study III, draft genomes passing INNUca QA/QC were annotated with Prokka (264), while pangenome analysis was carried out with Roary (105) (default parameters).

4.3.1 LOS locus annotation (I-III)

Glycosyltransferases were predicted based on the annotation available in the CAZy database (12). Assemblies were manually inspected with Artemis (261), and LOS locus classes were aligned and compared with ACT (265).

4.3.2 Improvement of *C. coli* 76339 LOS annotation (I)

To improve the gene prediction and to identify the transcription units for the genes included in the *C. coli* 76339 LOS biosynthesis locus a transcriptome analysis was performed. Total RNA was extracted from a culture grown for 16 h in nutrient broth 2 (Oxoid) (100 rpm, microaerobic atmosphere, and 37°C) using GeneJet RNA purification kit (Thermo Scientific). Dnase I treatment, rRNA depletion, and RNA sequencing were carried out by BaseClear (Netherlands). Single-end sequence reads were produced using the Illumina HiSeq2500 system. Illumina Casava pipeline version 1.8.3 was used to generate FASTQ sequence files. Adapters were clipped yielding reads of a minimum read length of 50 bp. The obtained FASTQ files were filtered for PhiX control signals and for reads that did not pass the Illumina Chastity. The final raw reads quality assessment was performed with FASTQC quality control tool version 0.10.0. Sequence reads were mapped against the *C. coli* 76339 genome with Rockhopper with default parameters (266).

4.4 Orthologue clustering and phylogenetic analysis (II)

A database including all *C. jejuni* and *C. coli* LOS biosynthesis translated coding sequences was assembled following Richards *et al.* orthologue nomenclature (17). Groups of orthologues (GOs) were determined with orthAgogue and MCL (103, 104), and aligned and

back-translated to nucleotide sequence with MUSCLE and Translatorx Perl script (267-269) as described in study II (270). Maximum likelihood phylogenetic reconstruction of each GO, in which *C. jejuni* and *C. coli* orthologues were admixed, was performed in MEGA6.06. To model the evolutionary rate differences among sites, the Kimura two-parameter model and a discrete gamma distribution (4 categories) was used (271). Results from 100 bootstrap runs were summarized in a 95% consensus tree.

4.5 Identification of *C. coli* GT-42 homologues (III)

C. coli reference genes homologous to *C. jejuni* GT-42 encoding genes were collected as follows. First, the non-redundant (nr) NCBI protein sequences collection was searched for *C. coli* Blastp hits (> 30% of amino acid identity; > 50% query coverage) using the amino acid sequences of CstI (Uniprot Q9RGF1), CstII (Uniprot Q9F0M9) and CstIII (Uniprot Q7BP25) and blast+ V 2.7.1 (272). Non-partial sequences were used for an all-versus-all blastp analysis, and classified into groups with > 0.7 Blastp Score Ratio (BSR) (273). A back-translated nucleotide sequence alignment of all the identified *C. coli* GT-42 proteins and *C. jejuni cstI*, *cstII,* and *cstIII* was built with MUSCLE (default parameters) (268). This alignment was used to infer a Minimum Evolution phylogenetic tree using MEGA (274).

4.6 Genome sequences mining, genes screening, and allele calling (III)

All the *Campylobacter coli* and *Campylobacter jejuni* whole genome raw sequence reads entries at the time of analysis (August 2017) were retrieved from the European Nucleotide Archive (ENA) using the ReMatCh framework v3.2 (https://github.com/B-UMMI/ReMatCh) (275).

To perform variant calling and inferring presence or absence of selected genes, reads were mapped against a set of reference genes (see below) using ReMatCh. A minimum coverage depth of 5x was required to consider a position to be present in the alignment, and 10x to perform allele calling. Additionally, to consider a locus to be present; successful mapping of at least 70% of the target reference gene sequence and ≥ 80% identity of the consensus sequence at nucleotide level were required.

4.6.1 *Campylobacter coli* clade identification

C. coli samples were assigned to one of the three previously described major phylogenetic clades (66, 276) by performing population structure analysis and inferred phylogenetic relationships based on the *atpA* gene (277). *C. coli* RM2228 *atpA* sequence of (KF855277) was used as reference for allele calling in all *C. coli* strains using ReMatCh as in 4.6. ReMatCh *atpA* consensus sequence alignment was then used for sample clustering at the first level with hierBAPS (278). MEGA7 (274) was then used to infer a Neighbor joining phylogenetic tree. To classify the clusters, *C. coli* strains representing the different clades

were used as reference (66, 276), and a group of *C. jejuni* strains were used as outgroup. The resulting tree was visualized in iTOL (279).

4.6.2 Frequency of GT-42 genes in *C. coli* and *C. jejuni*

To assess the frequency and distribution of GT-42 genes, identified *C. coli* GT-42 nucleotide sequences (see 4.5) were used as reference for calling orthologues in all *C. coli* and *C. jejuni* strains using ReMatCh as described in 4.6.

4.6.3 Classification of *C. coli* samples into a LOS locus class

To classify *C. coli* samples into a LOS class (17), orthologues in all GT-42 positive *C. coli* samples were called with ReMatCh, as described in 4.6, using the nucleotide sequences of loci located between the 'conserved putative two-domain glycosyltransferase' (orthologue 16 in Richards et *al.* (17)) and the 'LOS biosynthesis glycosyltransferase *waaV*' (orthologue 10 in Richards et *al.* (17)) from previously described *C. coli* LOS locus classes. Results were reported as a percentage of genes present for a given LOS locus class.

4.7 Pangenome analysis and gene flow investigation of LOS loci (III)

To infer HGT between *C. coli* clades, presence/absence of LOS associated group of orthologues were mapped into the *atpA* tree (see 4.6.1). To assess possible gene transfer between *C. coli* and *C. jejuni*, the nt NCBI database was searched with blastn using representative sequences of *C. coli* LOS associated group of orthologues. HGT between the two species was detected if the best blast hit for *C. jejuni* was > 90% nucleotide identity over > 70% of the *C. coli* query length.

4.8 Protein structure prediction

A protein sequence alignment of all characterized GT-42 enzymes was performed with MUSCLE (268). The sequence alignment was coloured with Sequence Manipulation Suite: Color Align Conservation (http://www.bioinformatics.org/sms2/color_align_cons.html). Three-dimensional structures of CstIV and CstV were predicted with the Phyre2 server (280). Both models were generated using *C. jejuni* CstII (1ro7) structure as a template. SuSPect was used to identify structurally, evolutionary, and functionally important regions (281). The quality of the generated models was assessed using the QMEAN Server for Model Quality Estimation (282).

4.9 LOS locus typing (II)

Previously reported LOS classes IV and V (17) were found to be the same LOS class by global alignment with progressiveMauve (283). Thus, they are considered as a single class, henceforth named IV/V. The LOS class of *C. coli* strains was determined by a newly

developed typing scheme (270). To determine the length of the LOS biosynthesis loci the region between orthologue 10 (LOS biosynthesis glycosyltransferase, *waaV*) and orthologue 16 (uncharacterized glycosyltransferase) (identifications numbers according to (17)) was amplified as follows: 25 µl reaction mixtures containing 0.5 U Phusion high-fidelity DNA polymerase (Thermo Scientific), 200 µM each deoxynucleoside triphosphate (dNTP) (Thermo Scientific), 0.4 µM each primer (ORF3F2 and waaV-R) (Table 1), 1× Phusion GC buffer (Thermo Scientific), 700 µM $MgCl_2$ (Thermo Scientific), and 50 ng of the template, were prepared. Then, PCR reactions were subjected to the following cycling conditions: 1 cycle at 98 °C for 30 s followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at 62.4 °C for 30 s, extension at 72 °C for 6 min, and a final elongation step at 72 °C for 6 min. PCR products were run in 0.5% SeaKem LE agarose in TBE buffer. Products were run at 60 V for 16 h. The 1-kb-plus (Thermo Scientific) and long-range (Thermo Scientific) molecular weight markers were used to estimate the size of the amplification products. To determine the putative LOS class, primers were designed based on specific features characterizing each LOS locus class using, when possible, multiple-sequence alignments of homologous sequences to improve sensitivity and specificity. Primer pairs and their amplicon sizes for each LOS class are shown in Table 1 of Study II, and a graphic representation of the primer annealing positions within the LOS locus is shown in Suppl. Fig. 2 of Study II. Preliminary gradient PCR was performed for each primer pair to select the most stringent conditions to minimize artefacts. PCRs were carried out in a semi-high-throughput manner; thus, isolates were classified into a LOS class based on the results of all PCRs. Isolates with an unexpected LOS size, negative for all tested orthologues, or with unexpected combinations of orthologues were classified as untypable.

4.10 LOS phenotype (II and IV)

LOS electrophoresis patterns were visualized by silver staining. To do so, the absorbance of the biomass obtained from a culture grown for 16 h in nutrient broth 2 (Oxoid) (100 rpm, microaerobic atmosphere, and 37 °C) was adjusted to an optical density of 0.5 at 600 nm (OD_{600}). Then, cells were digested with 20 mg/ml proteinase K (Thermo Scientific) at 55 °C for 1 h, followed by boiling for 10 min. Samples were then diluted 1:5 in SDS loading buffer and resolved in 15% SDS-PAGE gels. Gels were silver stained for visualization by fixing the gels in fixation solution (27% isopropanol, 7% acetic acid v/v) 24 h at with 40 rpm shaking. Then, gels were incubated in fixation solution with 7% periodic acid for 5 min and washed three times in ddH_2O for 15 min at 40 rpm. Afterwards, gels were stained in freshly prepared silver staining solution (0.87% $AgNO_3$, 0.52% NH_4OH in 0.024 M NaOH w/v) for 10 min at 70 rpm, followed by three washes in ddH_2O for 10 min at 40 rpm. Finally, gels were developed in a formaldehyde developer (0.93% formaldehyde in 1.3 mM citric acid, v/v). LOS sensitivity to neuraminidase was assessed by incubating crude LOS with 2 U/ml of *Clostridium perfringens* neuraminidase (Sigma) overnight at 37 °C. Electrophoretic patterns of neuraminidase treated *vs* untreated were visualized as above. The presence of Neu5Ac in *C. coli* 76339 LOS was assessed by high performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) as described in study

I (276). LOS chemical composition was determined by electrophoresis-assisted open tubular liquid chromatography–electrospray mass spectrometry (EA-OTLC-MS) analysis from lyophilized biomass as described in study II (270).

4.11 Construction of *C. coli* isogenic mutants (IV)

The Δ*cstIV*, Δ*cstV,* Δ*cstI,* Δ*neuB1,* Δ*neuB2,* and Δ*neuB3* mutants were constructed as follows. The genes *cstV* (orthologue 1501), *cstI* (BN865_06990), and *neuB1* (orthologue 9) in *C. coli* 76339, and Δ*cstIV* (orthologue 1501) in *C. coli* 65 and 73, and Δ*neuB2* in *C. coli* 73 were knocked out by homologous recombination with erythromycin resistance cassette (EryC) as described before (284), with the exception of *cstI* which was recombined with a chloramphenicol acetyltransferase cassette (CAT) (285). In brief, target genes with at least 300 bp downstream and upstream were amplified. Then, PCR fragments were restricted and cloned into pUC119 vector. Resulting plasmids were used as a template for inverse PCR. EryC or CAT were ligated to the inverse PCR product either in the same or opposite orientation as the gene of interest (Suppl. Table 1, Study IV). Generated plasmids were used as suicide plasmids (Suppl. Fig. 1-5, Study IV). Preparation of electrocompetent cells and transformations were done as previously described (285). Selection of Δ*cstIV*, Δ*cstV,* Δ*neuB1,* and Δ*neuB2* mutants was done on nutrient blood agar (NBA) supplemented with 10 μg ml^{-1} of erythromycin, while Δ*cstI* mutants were selected in NBA supplemented with 12.5 μg ml^{-1} of chloramphenicol. Homologous recombination of all mutants was verified by PCR. The complete list of primers and constructs generated can be found in the supplementary material of publication IV.

4.11.1 Complementation of *C. coli* 76339 ΔcstV-SR (IV)

Complementation of *C. coli* 76339 Δ*cstV*-SR4 was done in *cis* by integration of *cstV* under the active promoter of gamma glutalmyltranspeptidase (*ggt*). A suicide vector was built by amplifying *ggt* and ligating it to pGEM-T plasmid (Suppl. Fig. 6, Study IV). The resulting plasmid was subjected to inverse PCR and ligated to *cstV* and CAT. Homologous recombination of mutants was verified by PCR, RT-PCR and by testing for GGT activity as described earlier (177).

4.12 Sialyltransferase activity test in *C. coli* protein extracts (IV)

To test for sialyltransferase activity, *C. coli* 65, 73, and 76339 were grown for 16 h in nutrient broth 2 (Oxoid) (100 rpm, microaerobic atmosphere, and 37 °C). Cells were harvested by centrifugation and (10 000 × *g* for 15 min at 4°C) and resuspended in 50 mM HEPES pH 7.5 containing a protease inhibitor cocktail (Sigma). Cells were then lysed by mechanical disruption and debris was removed by centrifugation (10 000 × g for 15 min at 4°C).

Sialyltransferase activity of protein extracts was tested on boron-dipyrromethene or BODIPY (BDP) labelled Lac, LacNac, and 3'Siayllactose or fluorescein (FCHASE) labelled α-GalNAc, β-GalNAc, GM3, α-Gal, β-GlcNAc, α-Glc, β-Glc, Hep, as donors.

Reactions were performed at 37 °C in 10 µl volumes containing 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 1 mM CMP-NeuAc, 0.5 mM labelled acceptor, and 6 µl of extract. To stop enzymatic reactions, an equal volume of 80% acetonitrile was added. Enzymatic activity was assessed by thin-layer chromatography on silica using a solvent system of ethyl acetate/methanol/water/acetic acid 4:2:1:0.1.

4.13 Expression and activity of *C. coli* GT-42 enzymes (IV)

4.13.1 Construction of protein expression vectors

Gene *cstIV* from *C. coli* 73 and and *cstV* from *C. coli* 76339 were amplified and restricted. Both genes were ligated to pCw and pCw-MalET plamids (106) (See supplemental material from IV). Ligation products were then electroporated in *E. coli* 10β for plasmid amplification. After sequence confirmation, plasmids were electroporated into *E. coli* AD202 or *E. coli* BL21 for protein expression.

4.13.2 Protein expression, extraction, and screening

Cells containing the protein expression vectors were grown in 200 mL of 2YT medium supplemented with 150 µg/mL ampicillin and 0.2% of glucose at 25 °C with 250 rpm shaking. After reaching A_{600nm} of ~0.6, protein over-expression was induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and cultures were further incubated for 16 h. Cells were harvested by centrifugation (10 000 × *g* for 15 min at 4°C). Then, cells were resuspended and lysed as in 4.12. Crude protein extracts were run in 12% SDS-PAGE gels and stained with Coomassie blue to verify protein overexpression. Protein extracts were screened for sialyltransferase activity as in 4.12.

4.13.3 Cell free expression of CstIV and CstV

The genes *cstIV* and *cstV* were synthesized with a T7 promoter and a ribosome binding site upstream of the coding sequence and a T7 terminator downstream from the stop codon (Thermo Scientific). Synthesized products were inserted into pMA-T vector backbone. Plasmids were subjected to PCR with pExp-F and pExp-R primers. PCR products were verified in gel and purified with GeneJet PCR purification kit (Thermo Scientific). Purified PCR products were used as a template in cell-free *in vitro* translation with PURExpress *in vitro* Protein Synthesis Kit (New England Biolabs Inc.). Synthesized proteins were screened for sialyltransferase activity as in 4.12.

4.14 Statistical analysis (II and III)

Host-LOS locus class associations and clade and GT-42 associations were assessed by Fisher's exact test. *P* values of ≤0.05 were considered significant.

5 RESULTS

5.1 Novel *C. coli* LOS locus classes and GT-42 homologues (I-III)

Inspection of the first fully closed *C. coli* clade 3 genome (*C. coli* 76339) revealed a new LOS locus class; IX. LOS locus class IX contains four orthologues predicted to be involved in the synthesis and transfer of Neu5Ac: *neuA, neuB, neuC,* and *cstV*. The discovery of the first *C. coli* LOS locus class containing the essential genetic elements to potentially synthesize sialylated LOS structures prompted an in-depth study of *C. coli* LOS locus diversity.

The novel PCR typing scheme developed and validated in study II failed to classify 33.33% of the studied isolates into one of the seven previously reported *C. coli* LOS locus classes (17). Based on the sequencing results of eight untypable isolates, three new LOS locus classes were described: X, XI, and XII. In light of the elucidation of these three new *C. coli* LOS locus classes, the LOS typing scheme was expanded and all previously untypable strains were screened. Nevertheless, approximately 15% of the isolates remained untypable by the expanded PCR schema (Suppl. Fig. 2 in Study II).

Since no strain, aside from *C. coli* 76339, contained the genetic machinery to potentially express sialylated LOS, *C. coli* whole genome sequences available in public databases were systematically searched for the presence of GT-42 genes.

5.1.1 Identification of *C. coli* GT-42 homologues (III)

A group of *C. coli* GT-42 reference genes was collected by searching the non-redundant NCBI protein sequence collection using *C. jejuni* CstI, CstII, and CstIII (>30% amino acid identitiy; >50% query coverage). A total of 39 *C. coli* complete GT-42 coding sequences were identified. Sequences clustered into 7 different groups based on BSR, which varied between 0.8 and 0.98 (Table 1, Study III). These groups form monophyletic clades and are divided into two clusters: A and B (Fig. 1, Study III). Cluster A includes CstI, CstII, CstIII, and the newly identified CstVII, whereas cluster B comprises CstIV, CstV, and CstVI. In general, groups in cluster A were highly similar to *C. jejuni* GT-42 enzymes, while cluster B groups showed limited homology.

5.1.2 Frequency and distribution of *C. coli* GT-42 homologues (III)

To determine the frequency and distribution of GT-42 genes in the *C. coli* population a total of 2 432 *C. coli* genomes were mapped against all the 39 GT-42 reference sequences (See 5.1.1). Approximately 32% out of the genomes were positive for at least one GT-42 encoding gene (Table 2 and Suppl. Table 4, Study III). The *cstVI* was the most common GT-42 (49.35%), whereas *cstIII* was the rarest. Furthermore, ~5% of identified GT-42 genes were highly homologous to *C. jejuni* GT-42 enzymes. Statistical analysis revealed that *C.*

coli Clade 2 isolates are underrepresented among GT-42 positive genomes ($P < 0.0001$), while Clade 3 are overrepresented ($P < 0.001$). Additionally, 774 out of 2,432 (~32%) *C. coli* Clade 1 isolates were GT-42-positive. In *C. coli* Clade 1, *cstVI* (truncated) and *cstIV* were the most common. In contrast, *cstV* was the most common, and exclusively found, in *C. coli* Clade 3.

5.2 *C. coli* LOS locus classes containing GT-42 encoding genes (I and III)

Since gene content diversity is instrumental for LOS structural heterogeneity, genetic composition and organization of the LOS biosynthesis locus was investigated in GT-42 positive strains. An initial mapping of all GT-42 positive genomes against known *C. coli* LOS locus classes predicted that 99% of strains positive for *cstVI* have a LOS locus resembling class III (Suppl. Table S4, Study III). Also, a large percentage of strains positive for *cstV* (93%) and *cstIV* (68%) were predicted to possess a LOS locus similar to LOS classes IX and II, respectively. No significant resemblance to previously published *C. coli* LOS locus classes was found for genomes positive for other GT-42 genes. Consequently, all genomes positive for *cstI, cstII, cstIII,* or *cstVII*, and a subset of randomly selected *cstIV* and *cstVI* positive strains (261 genomes) were assembled and manually inspected to potentially identify novel LOS biosynthesis locus containing GT-42 genes. Apart from *cstI* and *cstVII*, all GT-42 genes are found within the LOS biosynthesis locus. A total of 27 different LOS locus classes were identified among the 160 assembled genomes found to contain a LOS-associated GT-42 gene. The group of LOS locus classes containing the *cstII* gene was the largest with 10 different gene content classes, followed by the *cstIV* (8 classes), *cstIII* (3 classes), *cstV* (3 classes), and *cstVI* (3 classes) groups (Fig. 2, Study III). Among the classes in the *cstIII* group, one was identical (100% homology) to *C. jejuni* LOS locus class C. Furthermore, sialic acid biosynthesis genes (*neuABC*) were present in LOS locus classes belonging to *cstII, cstIII,* and *cstV* LOS locus class groups, while they were absent in the *cstIV* and *cstVI* groups.

5.2.1 LOS locus and *C. coli* clade (II and III)

In *C. jejuni*, LOS locus classes are linked to population structure. To determine whether a similar phenomena is observed in *C. coli*, the distribution of GT-42 positive LOS locus classes between clades was investigated. A total of 11 LOS locus classes (II, III, XVI-XXIII, and *C. jejuni* LOS class C) were identified among 126 *C. coli* clade 1 isolates. LOS locus classes II and III were the most common accounting for 88.9% of *C. coli* Clade 1 isolates. Additionally, these LOS classes were found to be extremely conserved with ≥ 99% nt sequence identity over the whole locus. In contrast, 15 LOS locus classes (IX, XV, XXIII-XXXV) were represented among 32 *C. coli* Clade 3 isolates. The most common LOS locus class in *C. coli* clade 3, XXV, only represented 25% of the isolates. Overall, the LOS biosynthesis locus of *C. coli* Clade 1 was significantly less diverse ($P < 0.0001$) than that of Clade 3 strains. No conclusions on Clade 2 diversity could be drawn out due to the limited number of strains.

5.3 Flow of *C. coli* LOS genes (II and III)

The presence of LOS locus class XXIII in strains belonging to Clade 1 and Clade 3, and the high homology between *C. jejuni* and some *C. coli* GT-42 enzymes hinted at gene flow between *C. coli* clades and between species Therefore, gene flow between *C. coli* clades and between *C. coli* and *C. jejuni,* was assessed.

5.3.1 Gene flow between *C. coli* clades (III)

C. coli isolates belonging to different clades were found to share LOS-associated orthologues according to orthologue group delineation by Roary (>95% amino acid identity) (Fig. 3, Study III). The vast majority of the shared orthologues were predicted to encode proteins putatively involved in sugar biosynthesis or sugar modification (Table 3, Study III). Also, *cstII* and *cstVI* orthologues were found to be shared between *C. coli* Clade 1 and Clade 3.

5.3.2 Gene flow between *C. coli* and *C. jejuni* (II and III)

The impact of introgression between *C. jejuni* and *C. coli* on the LOS biosynthesis locus was investigated. In previously published LOS locus classes, comparison of individual gene descriptions and subsequent analysis of tree topology indicated that interspecies gene transfer had only occurred in *waaV* (Fig. 2, Study II). However, amidst the identification of novel *C. coli* LOS locus classes, gene flow from *C. jejuni* to *C. coli* between the two species was detected in various instances. First, a high sequence similarity was observed between *C. coli* and *C. jejuni cstII*-associated LOS locus classes. Second, *C. coli* LOS classes XVI, XVII, XVIII, XIX, XX, XXI, and XXII were found to be a mosaic of *C. jejuni* LOS classes A, B, S/F, and I/D. Third, *C. coli* LOS class XXIV contains GT42 and *neuB1* orthologues characteristic of *C. coli* and *C. jejuni* (i.e. *cstV*, *cstII*). Finally, a *C. coli* strain was found to possess a *C. jejuni* LOS locus class C, evidencing cross-species mobilization of the entire LOS biosynthesis locus.

To assess whether gene flow of GT-42 genes was monodirectional (*C. jejuni*→*C. coli*) or bidirectional (*C. jejuni*↔ *C. coli*), a total of 12 391 raw genome sequences deposited as *C. jejuni* were mapped against the 7 *C. coli* GT-42 groups. Of the 7 577 putative *C. jejuni* genomes positive for at least one GT-42 gene, 95.75% contained either a *cstII* or a *cstIII*. Genes *cstIV* (101 genomes), *cstVI* (211), and *cstVII* (2) were present in a small fraction of the population, while *cstV* was not detected. To inspect the LOS locus composition of isolates positive for *cstIV, cstVI,* and *cstVII*, genomes were assembled and the species was verified. Surprisingly, solely 16.6% of the genomes were confirmed as *C. jejuni,* the rest were either *C. coli* or a mixture of both species. Among the verified *C. jejuni*; 40 were positive to *cstIV*, 10 to *cstVI,* and 2 to *cstVII.* Like in *C. coli*, *cstIV* and *cstVI* were LOS associated, while *cstVII* was located outside the LOS locus. Additionally, the LOS loci of 87.5% of the *cstIV* positive *C. jejuni* genomes were 99% identical to *C. coli* LOS class II, demonstrating the occurrence of gene flow from *C. coli* to *C. jejuni*.

5.4. Origin of LOS locus classes II and III (III)

Recognition of gene flow between *C. coli* Clades, and the apparent absence of other LOS locus classes containing *cstIV* or *cstVI* in *C. coli* Clade 1 LOS, incited a search for the origin of the most abundant LOS locus classes in *C. coli* Clade 1; II and III. LOS locus class II exhibited ~88% nucleotide identity over ~99% of length with LOS locus class XXXIV found in Clade 3 (Fig. 4).

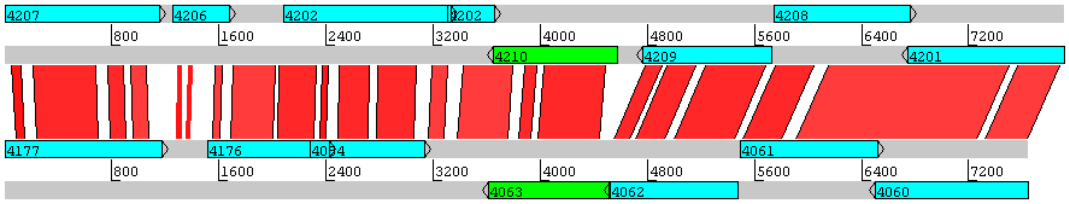

Fig. 4 ACT comparison LOS locus class II (top) vs LOS locus class XXXIV (bottom).

Similarly, LOS locus class III shares high sequence similarity (>90%) with LOS locus classes XXVII and XXVIII found in Clade 3 (Fig. 5 and 6) over the terminal part of the locus.

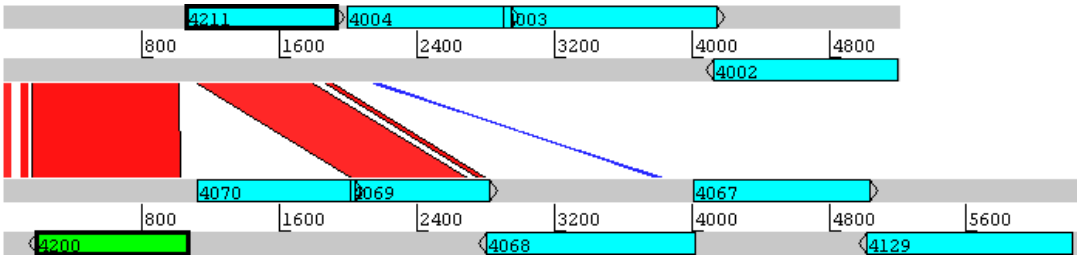

Fig. 5 ACT comparison LOS locus class III (top) vs LOS locus class XXVII (bottom)

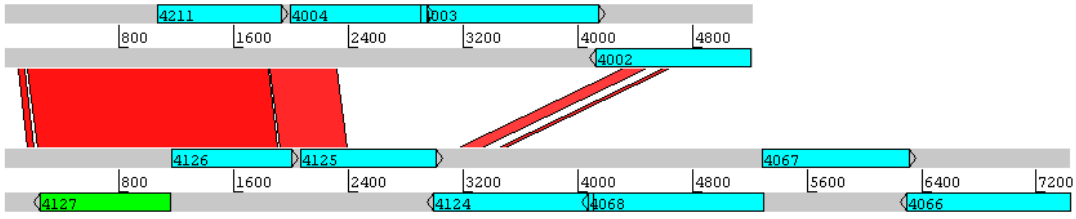

Fig. 6 ACT comparison LOS locus class III (top) vs LOS locus class XXVIII (bottom)

Interestingly, while the putative phosphoethanolamine transferase genes (*eptC*) and the *cstVI* in Clade 3 encode full proteins, these genes underwent pseudogenization in Clade 1.

Thus, implying that LOS locus classes II and III presumptively originated in *C. coli* Clade 3.

5.5 LOS locus class groups and source of isolation (II and III)

As results suggested that *C. coli* LOS locus classes II and III might have undergone diversification as a consequence of adaptation to the agricultural niche, the distribution of GT-42 positive LOS locus class groups by source of isolation was investigated in isolates for which metadata was available. Results are shown on Figure 7.

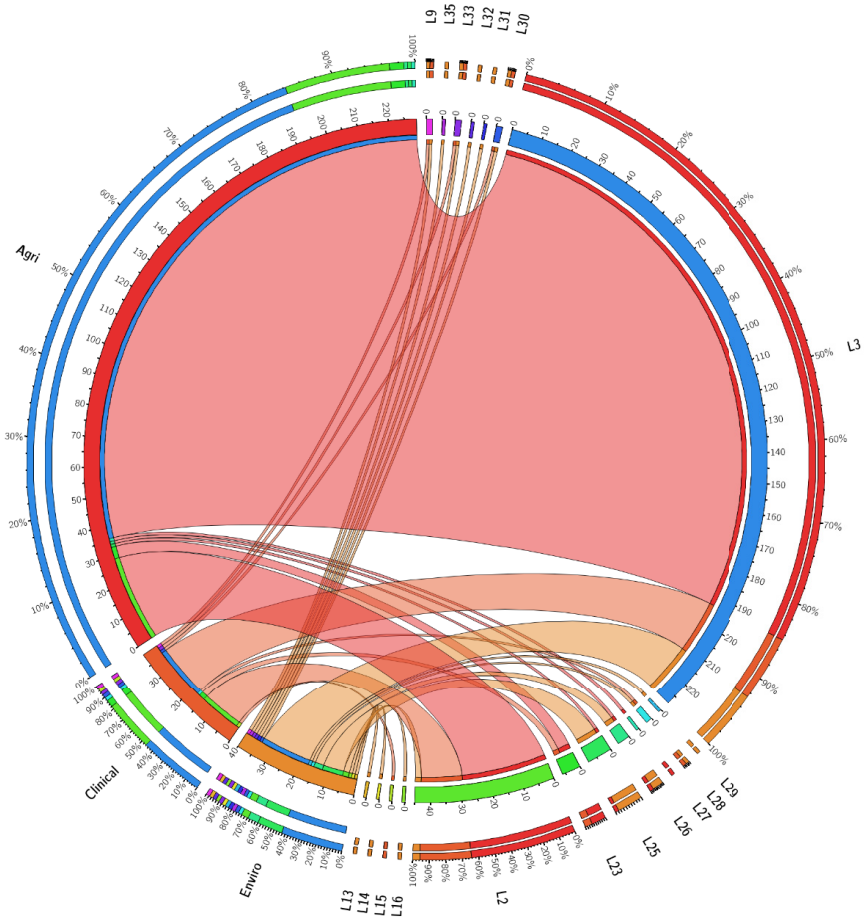

Figure 7. Source-LOS locus class association. Circus diagram showing the distribution of LOS locus classes of *C. coli* strains isolated from different sources. Ribbon ends represent links between source and LOS locus class, while the width of the ribbon correlates with percentage of strains belonging to a certain LOS locus class from a particular source. Segments in the outer ring indicate the percentage of strains representing a certain LOS locus class or source of isolation, while the inner ring indicates the number of strains. Agricultural strains (Agri) are shown in red, clinical strains in dark orange, and environmental strains (Enviro) in light orange.

As expected, LOS locus classes II and III were dominant among agricultural isolates. Conversely, a small proportion of strains carrying a LOS locus class II (4.4%) or III (7.5%) were isolated from the environment. Although the number of strains was small, other LOS locus classes containing either *cstIV* or *cstVI* were not found in agricultural isolates. Seven strains carrying LOS locus classes XXIII, XXV XXVI, or XXVII, were also isolated from agricultural settings. Surprisingly, three of these strains belonged to *C. coli* Clade 3.

5.6 CstIV and CstV (I, II, IV)

The link between LOS-associated GT-42 encoding genes and LOS sialylation has only been established for *cstII* and *cstIII,* so far. However, the presence of other GT-42 genes among clinical isolates instigated research on the potential association between these genes and LOS structure. As *cstVI* is generally found as a pseudogene, efforts were focused on elucidating the role of *cstIV* and *cstV* in LOS biosynthesis. *C. coli* 65 and 73 (LOS locus class II, *cstIV*) and *C. coli* 76339 (LOS locus class IX, *cstV*) were selected as model strains

5.6.1 Enzymatic activity (IV)

Given that all GT-42 proteins characterized to date are sialyltransferases, the ability of CstIV and CstV to transferring Neu5Ac was tested. Crude protein extracts of *C. coli* 65, 73, and 76339 were tested for sialyltransferase activity *in vitro* using BDP or FCHASE labelled sugar acceptors. *C. coli* 76339 WT, *C. coli* 76339 Δ*cstV*-SF1, and *C. coli* 76339 Δ*cstV*-SR4 crude protein extracts exhibited monosialyltransferase activity on BDP-Lactose (BDP-Lac) and BDP-*N*-acetyllactosamine (BDP-LacNAc) (Suppl. Fig. 1, Study IV). However, sialyltransferase activity was undetectable on any of the tested acceptors in the crude protein extracts of *C. coli* 76339Δ*cstI*-XR3 and *C. coli* 76339Δ*cstV*-SRΔ*cstI*-XR1. Therefore detected monosyltransferase activity in *C. coli* 76339 WT protein extracts was due to presence of CstI. Similarly, no sialyltransferase activity was detected in the crude protein extracts of *C. coli* 65 and 73, *C. coli* 65 Δ*cstIV*-SF5 and *C. coli* 73 Δ*cstIV*-SF3.

Since *cstIV* and *cstV* were confirmed to be expressed in the tested strains, it was hypothesized that the concentration of CstIV and CstV in the protein extracts was too low to be able to detect sialyltransferase activity. Consequently, CstIV and CstV were produced as recombinant proteins in *E. coli* and tested for sialyltransferase activity. Despite different strategies for increasing protein solubility, facilitating disulphide formation, encouraging protein folding, and increasing overall protein yield, none of the tested recombinant CstIV and CstV showed activity on any of the tested acceptors.

5.6.2 Role in *C. coli* LOS biosynthesis (I and IV)

Since the LOS of *C. coli* 65, 73, and 76339 was insensitive to neuraminidase, potentially suggesting the absence of Neu5Ac (Suppl. Fig. 4, Study IV), mutational studies were performed to determine whether CstIV and CstV were involved in LOS biosynthesis. Deletion of *cstV* in *C. coli* 76339 and *cstIV* in *C. coli* 65 and 73 resulted in LOS truncation (Fig. 1, Study IV). This implied that *cstV* in *C. coli* 76339, and *cstIV*, in *C. coli* 65 and 73, are involved in LOS biosynthesis.

To investigate whether CMP-Neu5Ac is CstV substrate, *neuB1* was knocked out. The LOS of *cstV* and *neuB1* mutants showed a similar truncation, hence, suggesting the potential involvement of *neuB1* in CstV substrate synthesis. Analysis by HPAEC-PAD and LC-MS detected the presence of Neu5Ac in *C. coli* 76339 LOS. Furthermore, comparison of the predicted LOS compositions of *C. coli* 76339 WT and the *cstV* mutant showed the loss of Neu5Ac. Thus, establishing the association between *cstV* and NeuAc.

5.6.3 Exploring CstIV substrate (II-IV)

While the presence of Neu5Ac was confirmed in *C. coli* 76339, no evidence of sialylation was found in *C. coli* 65 and 73 by EA-OTLC-MS. However, further analysis by LC-MS suggested the presence of a diNAc-nonulosonate with a mass consistent with that of Leg5Ac7Ac or Pse5Ac7Ac. Comparison of the predicted LOS compositions of *C. coli* 73 WT and the *cstIV* mutant showed the loss of the diNAc-nonulosonate residue. Hence, implying that *cstIV* is associated to the diNAc-nonulosonate residue. LOS locus class II does not contain any homologues known to be involved in nonulosonic acid biosynthesis. Additionally, genome-wise analysis did not identify genes overrepresented in LOS locus class II strains that might potentially be involved carbohydrate biosynthesis. However, legionaminic (*neuB2*) and pseudaminic (*neuB3*) synthetases are found within the conserved flagella glycosylation system. Thus, the role of these genes in the synthesis of CstIV substrate was investigated. No difference in the electrophoretic profiles *C. coli* 73 WT and *C. coli* 73Δ*neuB2* LOS was observed, implying that *neuB2* is not associated to the synthesis of CstIV substrate. Despite repeated attempts, no *C. coli* 73Δ*neuB3* mutants were obtained. This may suggest that deletion of *neuB3* is lethal for *C. coli* 73.

6 DISCUSSION

As a result of being major cell surface components, glycoconjugates have a considerable impact on host-pathogen interactions, and thus in the ecology and evolution dynamics of bacteria (4, 5, 92, 94, 95). Expression of host-like glycans generally facilitates evasion of the human immune system by pathogens (214, 286). However, recognition of *C. jejuni* ganglioside-like LOS as an antigen may result in the production of self-reacting antibodies and development of GBS (11, 33-35, 38, 40, 41)

Approximately 58% of *C. jejuni* genomes available in public repositories (at the time of analysis, **Study III**) contain genes associated to biosynthesis and transfer of sialic acid to the LOS chain (186, 230, 286). Despite being isolated from GBS patients in few instances (13-16) and the pervasive introgression with *C. jejuni*, no *C. coli* containing the required molecular machinery for LOS sialylation has been described until now (17). The clinical strain 76339 was the first *C. coli* reported to possess a LOS locus containing a GT-42 gene (*cstV*) upstream of sialic acid biosynthesis genes. Thus, *C. coli* 76339 LOS locus partially resembled *C. jejuni* LOS locus classes A and B (**Study I**) (17, 175, 186). Typing of a large strain collection indicated that *C. coli* 76339 LOS locus was rare (**Study II**). Yet, a systematic search of all publicly available *C. coli* genomes showed that the *cstVneuBCA* gene cluster is restricted to LOS locus classes exclusively detected in *C. coli* Clade 3 strains (**Study III**). Thus, the apparent singularity of *C. coli* 76339 LOS locus was due to the overrepresentation of agricultural strains from Clade 1 in the isolate collection (**Study II**).

In addition to *cstV,* six novel GT-42 homologues were identified in the study herein (**Study I and III**), showing that *C. coli* possess a larger GT-42 enzyme repertoire than *C. jejuni* (17, 175, 186). Only two GT-42 orthologues, *cstI* and *cstVII,* are located outside the LOS biosynthesis locus in all the studied genomes (**Study III**). As previously observed for *C. jejuni cstI*, GT-42 genes located outside the LOS biosynthesis locus are unlikely to have an impact in LOS biosynthesis (96). Supporting this assumption, deletion of *cstI* in *C. coli* 76339 had no effect on LOS biosynthesis (**Study IV**).

Although glycan structure is not template driven, *C. jejuni* strains carrying LOS associated *neuABC* and *cstII* or *cstIII* have been shown to rarely synthesize non-ganglioside sialylated LOS (7-10). Thus, these genes are utilized as markers for recognizing *C. jejuni* strains potentially expressing ganglioside mimics (9, 10, 179). The presence of *neuABC* genes in LOS locus classes containing *cstII* and *cstIII,* hinted at the possibility of ganglioside mimicry in *C. coli* (**Study III**). Backing this notion, a complete *C. jejuni* LOS class C was discovered in *C. coli* strain SRR5152313 (100% homology). Hence, this strain may synthesize a GM1a- or GM2-like LOS (9).

Originally hypothesized to be unaffected by introgression (66), phylogenetic analysis of previously published LOS locus genes confirmed limited genetic exchange between *C. jejuni* and *C. coli* (**Study II**). Among the identified shared orthologues, only *waaV* in LOS locus class II exhibited some evidence of recombination. However, the presence of *cstII* and

cstIII in the LOS biosynthesis locus of various *C. coli* strains implied that some LOS locus classes have arisen as consequence of introgression. Although further studies are required, results obtained in **Study II and III** indicate that only GT-42 containing LOS locus classes appear to be subjected to extensive recombination. In fact, all 10 *C. coli* LOS locus classes containing a *cstII* are mosaics of *C. jejuni* LOS classes (**Study III**). This indicates that genetic factors associated to ganglioside mimicry can cross species barriers. Additionally, the presence of *cstII* containing LOS locus classes XV, XXIII, and XXIV in *C. coli* Clade 3 isolates points at the potential transfer of these factors across clades.

Given the number of strains found to carry a *C. jejuni*-like LOS, it appears that the frequency at which HGT occurs is low. Overall, strains likely to express sialylated LOS are a minority, 1.8%, among sequenced *C. coli* strains. Yet, this seeming rarity might be a consequence, in part, of sampling bias. Despite the relative low frequency, these data clearly suggest that the participation of *C. coli* in the GBS aetiology cannot be excluded. Moreover, it is worth noticing that even though only 3.2% of the currently known *C. coli* genomes belong to Clade 3, the molecular machinery involved in the synthesis of sialylated LOS structures (i.e. *neuABC* and GT-42 genes) is strongly associated with this lineage. In contrast, the sialic acid biosynthesis locus is rarely found in Clade 1 GT-42 positive strains.

The genes *cstIV* and *cstVI* are the most common GT-42 encoding genes among sequenced *C. coli* strains. Additionally, these GT-42 genes are exclusively found in LOS locus classes lacking the sialic acid biosynthesis genes: II and III, respectively. The LOS locus classes II and III are the most common LOS locus classes among sequenced *C. coli* strains. Despite being most likely originated from non-agriculture Clade 3 LOS classes, LOS locus classes II and III are restricted to strains belonging to the agricultural-adapted Clade 1 (**Study III**). During the process of evolution from Clade 3, in both LOS class II and III, various genes have undergone pseudogenization. Loss of activity, such as in the case of *eptC* in class II or *cstVI* in class III, may have impacted LOS structure and composition. Diversification of surface structures as a consequence of natural selection is common in microbial evolution. For instance, antigenic shift is a successful mechanism through which *Neisseria meningitidis* circumvents vaccine-induced protective immunity. As a result, immunization campaigns may select for capsular switching to serogroups not cover by the polysaccharide vaccine (287-289). Similarly, structural variability in *C. jejuni* capsule influence phage infectivity. Thus, phage predation in the avian gut has resulted in the so-call 'Red Queen' effect; evasion of rapidly evolving pathogens by altering glycan expression, without compromising survival (290, 291).

The pseudogenization observed in the two most frequently detected LOS locus classes in Clade 1 might be the result of the recent clonal expansion of agriculture adapted population (**Study III**). In fact, *C. coli* clade evolution has been shaped by the introduction of the agricultural niche (64). A plausible evolutionary scenario for the LOS locus classes containing *cstIV* or *cstVI* is that adaptation to the agricultural niche caused a change in selective pressure. As a result, glycan structures associated to LOS locus classes II and III

might have provided a competitive advantage. This scenario is supported by three observations: (i) the predominance of LOS locus classes II and III in *C. coli* Clade 1 strains, (ii) the limited nucleotide variability in these LOS locus classes (>99% identity), hinting a strong purifying selection, and (iii) the presumable absence of LOS locus classes II and III in the environmentally associated *C. coli* Clades 2 and 3. Moreover, the hypothesis of LOS locus classes II and III playing a key role in adaptation to the agricultural niche is also backed by the presence of these genetic elements in the bovine associated *C. jejuni* ST-459 (292). Nevertheless, further investigation is required to ascertain the role of structures associated to LOS locus classes II and III in niche adaptation.

Adaptation of *C. coli* Clade 1 to specific animal niches might be linked to the high-affinity interactions between the oligosaccharide moiety in the LOS and host glycans, as previously suggested (5). Indeed, distribution of *C. coli* LOS locus classes groups among sources of isolation vastly reflects strain affiliation (**Study II and III**; (17)). For example, strains possessing non-GT-42 LOS locus classes VI and X were predominantly isolated from pigs, which is known to be frequently colonized by *C. coli* (**Study II;** (293)). However, both of these classes are rarely detected in humans. In the case of clinical isolates, the GT-42 containing LOS locus classes II and III were the most common (**Study II**). These classes were also detected in multiple animal species (i.e. porcine, poultry and bovine) (**Study II**), suggesting the generalist nature of strains possessing these features and their importance in human disease.

Given their potential significance for the pathogenicity and ecology of *C. coli,* the role of CstV (class IX) and CstIV (class II) in LOS biosynthesis was further explored (**Study IV**). Although the presence of Neu5Ac was initially detected in *C. coli* 76339 purified LOS by HPAEC-PAD (**Study I**), the occurrence of a secondary active monosyltransferase, CstI (**Study IV**), could not rule out the possibility of contamination with a non-LOS sialylated glycan.

Identical LOS electrophoretic profiles were observed upon deletion of *neuB* or *cstV*, suggesting a similar truncation in the LOS structure, and the possible involvement of *neuB* in the synthesis of CstV substrate. According to LC-MS analysis, deletion of *cstV* results in the loss 2 Hex and 1 NeuAc. *C. coli* 76339 LOS is neuraminidase resistant, implying that a Neu5Ac is internally located in the LOS, as seen in *C. jejuni* 11168 (206). Consequently, absence of Neu5Ac could potentially impede further extension of the oligosaccharide chain.

Attempts at confirming CstV activity *in vitro* were unsuccessful. Transcriptomic analysis of *C. coli* 76339 revealed the expression of LOS biosynthesis genes in a polycistronic fashion, confirming the active expression of *cstV*. However, upon deletion of *cstI,* no sialyltransferase activity was detected in *C: coli* 76339 protein extracts. Thus, it was hypothesized that the absence of detectable activity was a consequence of a low concentration of CstV in *C. coli* 76339 protein extracts. Despite numerous efforts and various protein expression strategies, no detectable activity on any of the tested acceptors was exhibited by recombinant CstV. Taken together, *cstV* is associated to *C. coli* 76339

LOS sialylation. However, whether *cstV* is involved in the synthesis of human-like glycans remains an open question. As a result, further studies are required to identified CstV natural acceptor and corroborate its activity *in vitro.*

An initial analysis of crude LOS samples from various strains expressing *cstIV* (LOS class II) failed to detect Neu5Ac (**Study II**). LOS asialylation was concordant with rare presence of Neu5Ac biosynthesis genes in strains carrying *cstIV* (**Study III**). Like in *cstV,* no sialyltransferase activity was detected in the protein extracts of two LOS locus class II strains (*C. coli* 65 and 73) nor in recombinant CstIV despite several attempts. Consequently, CstIV appeared to be a cryptic, possibly inactive, sialyltransferase with no role in LOS biosynthesis. Surprisingly, deletion of *cstIV* in *C. coli* 65 and 73 resulted in LOS truncation. Analysis of CstIV amino acid sequence revealed numerous substitutions at conserved positions in characterized GT42 sialyltransferases. Moreover, substitutions of amino acids involved in substrate interactions, particularly with the Neu5Ac, were observed upon superimposition of CstIV on *C. jejuni* CstII (118, 119, 232, 294). Alignment of multiple CstIV confirmed the conservation of these substitutions. Based on these results, the possibility of an alternative sugar donor for CstIV was considered. While no insights into genes linked to the synthesis of CstIV substrate were gained by genome-wise analysis (**Study III**), detection of a diNAc-nonulosonate residue in *C. coli* 73 WT LOS and its absence in the *cstIV* mutant suggested *neuB2* (*ptmC, legI*) and *neuB3* (*pseI*) as potential candidates. The *neuB2* (*ptmC, legI*) and *neuB3* (*pseI*) encodes a legionaminic acid and a pseudaminic acid synthetase, respectively. In *C. coli,* these conserved genes are associated to the post-translational modification of flagellin with nonulosonic acids (131, 136-138, 211, 230, 295). *C. coli* 73 Δ*neuB2* expressed a LOS with a similar molecular weight as the WT, implying that *neuB2* has no role in LOS biosynthesis. While deletion of *neuB3* has been successful in *C. coli* VC167, no viable *C. coli* 73 Δ*neuB3* mutants were obtained despite repeated attempts (137). It is hypothesized that flagellin glycosylation and perhaps LOS truncation may be lethal for *C. coli* 73. In summary, *cstIV* is associated with the presence of a diNAc-nonulosonate residue in *C. coli* 73 LOS. Although further studies are required to determine the nature of this residue, these results highlight the importance of functional studies, as divergent evolution can result in genes encoding proteins with differing activities (**Study I**).

7 CONCLUSIONS

1. *C. coli* LOS locus classes are more diverse than previously thought. The 35 *C. coli* LOS locus classes, including the 26 identified in this study, represent a fraction of the LOS locus classes found in the *C. coli* population. Further identification of novel LOS locus classes will deem the current LOS class classification system results impractical. Grouping of LOS locus classes by their association to distinctive structural features, for example Neu5Ac or Qui3NAcyl would result more valuable. However, this would require extensive functional and structural studies.
2. Marked differences in the genetic diversity of the LOS locus was observed between *C. coli* Clade 1 and *C. coli* Clade 3. The vast majority of *C. coli* Clade 3 strains possessed a unique LOS locus class. Additionally, these classes generally contain the required genetic machinery for the expression of sialylated LOS structures. Conversely, over 80% of *C. coli* clade 1 strains possess a LOS locus class II or III, which lack the sialic acid biosynthesis genes. As a result, the LOS of *C. coli* Clade 1 strains is rarely sialylated. Overall, the observed genetic differences are possibly a reflection of host adaptation.
3. LOS genes can cross clade and species barriers. Several orthologues common to all *C. coli* clades were identified. More importantly, bacterial factors implicated in GBS aetiology have been acquired from *C. jejuni* and can potentially spread across the *C. coli* population, as demonstrated by the presence of *cstII* and *cstIII* in all three *C. coli* clades. Consequently, participation of *C. coli* in the GBS aetiology cannot be excluded.
4. *C. coli* presents a larger GT-42 enzyme repertoire than *C. jejuni*. CstV is a native *C. coli* Clade 3 sialyltransferase, while CstIV is associated to the addition of a diNAc-nonulosonate residue to the LOS chain. Based on the frequency of these genes in *C. coli* population, nonulosonic acids in *C. coli* LOS appear to be a common feature.

8 REFERENCES

1. **Blow, N.** 2009. Glycobiology: A spoonful of sugar. Nature. **457:**617-620. http://dx.doi.org/10.1038/457617a.

2. **Bergman, M., G. Del Prete, Y. van Kooyk, and B. Appelmelk.** 2006. *Helicobacter pylori* phase variation, immune modulation and gastric autoimmunity. Nat Rev Micro. **4:**151.

3. **WHO.** 2015. **WHO estimates of the global burden of foodborne diseases: foodborne diseases burden epidemiology reference group 2007-2015**. .

4. **Day, C. J., E. A. Semchenko, and V. Korolik.** 2012. Glycoconjugates play a key role in *Campylobacter jejuni* Infection: interactions between host and pathogen. Front Cell Infect Microbiol. **2:**9. doi: 10.3389/fcimb.2012.00009 [doi].

5. **Day, C. J., E. N. Tran, E. A. Semchenko, G. Tram, L. E. Hartley-Tassell, P. S. Ng, R. M. King, R. Ulanovsky, S. McAtamney, M. A. Apicella, J. Tiralongo, R. Morona, V. Korolik, and M. P. Jennings.** 2015. Glycan:glycan interactions: High affinity biomolecular interactions that can mediate binding of pathogenic bacteria to host cells. Proc Natl Acad Sci USA. **112:**E7266-75. doi: 10.1073/pnas.1421082112 [doi].

6. **Parkhill, J., B. W. Wren, K. Mungall, J. M. Ketley, C. Churcher, D. Basham, T. Chillingworth, R. M. Davies, T. Feltwell, S. Holroyd, K. Jagels, A. V. Karlyshev, S. Moule, M. J. Pallen, C. W. Penn, M. A. Quail, M. Rajandream, K. M. Rutherford, van Vliet, A. H. M., S. Whitehead, and B. G. Barrell.** 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature. **403:**665-668. http://dx.doi.org/10.1038/35001088.

7. **Gilbert, M., M. Karwaski, S. Bernatchez, N. M. Young, E. Taboada, J. Michniewicz, A. Cunningham, and W. W. Wakarchuk.** 2002. The genetic bases for the variation in the lipo-oligosaccharide of the mucosal pathogen, *Campylobacter jejuni*. J. Biol. Chem. **277:**327-337. doi: 10.1074/jbc.M108452200.

8. **Godschalk, P. C. R., M. L. Kuijf, J. Li, F. St. Michael, C. W. Ang, B. C. Jacobs, M. Karwaski, D. Brochu, A. Moterassed, H. P. Endtz, A. van Belkum, and M. Gilbert.** 2007. Structural characterization of *Campylobacter jejuni* lipooligosaccharide outer cores associated with Guillain-Barré and Miller Fisher syndromes. Infect. Immun. **75:**1245-1254. doi: 10.1128/IAI.00872-06.

9. **Godschalk, P., A. Heikema, and M. e. a. Gilbert.** 2004. The crucial role of *Campylobacter jejuni* genes in anti-ganglioside antibody induction in Guillain-Barré syndrome. J. Clin. Invest. **114:**1659-1665.

10. **Houliston, R. S., E. Vinogradov, M. Dzieciatkowska, J. Li, F. St. Michael, M. Karwaski, D. Brochu, H. C. Jarrell, C. T. Parker, N. Yuki, R. E. Mandrell, and M. Gilbert.** 2011. Lipooligosaccharide of *Campylobacter jejuni*: similarity with multiple

types pf mammalian glycans beyond gangliosides. J Biol Chem. **286:**12361-12370. doi: 10.1074/jbc.M110.181750.

11. **Yuki, N.** 2005. Carbohydrate mimicry: a new paradigm of autoimmune diseases. Curr. Opin. Immunol. **17:**577-582. doi: http://dx.doi.org/10.1016/j.coi.2005.09.004.

12. **Lombard, V., H. Golaconda Ramulu, E. Drula, P. M. Coutinho, and B. Henrissat.** 2013. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. **42:**D490-D495. doi: 10.1093/nar/gkt1178. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3965031/.

13. **Wulffen, H. v., C. Hartard, and E. Scharein.** 1994. Seroreactivity to *Campylobacter jejuni* and gangliosides in patients with Guillain-Barré syndrome. J. Infect. Dis. **170:**828-833.

14. **Funakoshi, K., M. Koga, M. Takahashi, K. Hirata, and N. Yuki.** 2006. *Campylobacter coli* enteritis and Guillain–Barre syndrome: No evidence of molecular mimicry and serological relationship. Journal of the Neurological Sciences. **246:**163-168.

15. **Bersudskya, M., P. Rosenbergb, B. Rudenskyc, and I. Wirguin.** 2000. Lipopolysaccharides of a *Campylobacter coli* isolate from a patient with Guillain-Barre syndrome display ganglioside mimicry. Neuromuscular Disorders. **10:**182-186.

16. **van Belkum, A., B. Jacobs, E. van Beek, R. Louwen, W. van Rijs, L. Debruyne, M. Gilbert, J. Li, A. Jansz, F. MÃ©graud, and H. Endtz.** 2009. Can *Campylobacter coli* induce Guillain-Barré syndrome? European Journal of Clinical Microbiology & Infectious Diseases. **28:**557-560. doi: 10.1007/s10096-008-0661-9. http://dx.doi.org/10.1007/s10096-008-0661-9.

17. **Richards, V. P., T. Lefébure, P. D. Pavinski Bitar, and M. J. Stanhope.** 2013. Comparative characterization of the virulence gene clusters (lipooligosaccharide [LOS] and capsular polysaccharide [CPS]) for *Campylobacter coli, Campylobacter jejuni* subsp. *jejuni* and related *Campylobacte*r species. Infect Genet Evol. **14:**200-213. doi: 10.1016/j.meegid.2012.12.010.

18. **Naess, V., and T. Hofstad.** 1984. Chemical studies of partially hydrolysed lipopolysaccharides from four strains of *Campylobacter jejuni* and two strains of *Campylobacter coli*. J. Gen. Microbiol. **130:**2783-2789. http://europepmc.org/abstract/MED/6527123.

19. **Beer, W., M. Adam, and G. Seltmann.** 1986. Monosaccharide composition of lipopolysaccharides from *Campylobacter jejuni* and *Campylobacter coli*. J. Basic Microbiol. **26:**201-204. doi: 10.1002/jobm.3620260405. http://europepmc.org/abstract/MED/3772734; http://dx.doi.org/10.1002/jobm.3620260405.

20. **Naess, V., and T. Hofstad.** 1984. Chemical composition and biological activity of lipopolysaccharides prepared from type strains of *Campylobacter jejuni* and

Campolybacter coli. Acta Path Microbiol Immunol Scand. **92B:**217-222. doi: 10.1111/j.1699-0463.1984.tb02824.x.

21. **Aspinall, G. O., A. G. McDonald, H. Pang, L. A. Kurjanczyk, and J. L. Penner.** 1993. Lipopolysaccharide of *Campylobacter coli* serotype O:30. Fractionation and structure of liberated core oligosaccharide. J. Biol. Chem. **268:**6263-6268.

22. **Seltmann, G., and W. Beer.** 1985. Vorkommen von 3-Amino-3,6-didesoxy-D-glucose in einem Lipopolysaccharid von *Campylobacter coli* . J. Basic Microbiol. **25:**551-552.

23. **Skirrow, M. B.** 1977. *Campylobacter* enteritis: a "new" disease. British Medical Journal. **2:**9-11.

24. **Ketley, J. M.** 1997. Pathogenesis of enteric infection by *Campylobacter*. 143. 5-21.

25. **Hughes, R. A., and D. R. Cornblath.** 2005. Guillain-Barré syndrome. The Lancet. **366:**1653-1666. doi: http://dx.doi.org/10.1016/S0140-6736(05)67665-9.

26. **Sejvar, J. J., A. L. Baughman, M. Wise, and O. W. Morgan.** 2011. Population incidence of Guillain-Barré syndrome: a systematic review and meta-analysis. Neuroepidemiology. **36:**123-133. http://www.karger.com/DOI/10.1159/000324710.

27. **Huizinga, R., B. van den Berg, W. van Rijs, A. P. Tio-Gillen, W. J. R. Fokkink, L. E. Bakker-Jonges, K. Geleijns, J. N. Samsom, P. A. van Doorn, J. D. Laman, and B. C. Jacobs.** 2015. Innate immunity to *Campylobacter jejuni* in Guillain-Barré syndrome. Ann. Neurol. **78:**343-354. doi: 10.1002/ana.24442.

28. **Nobile-Orazio, E., M. Carpo, N. Meucci, M. P. Grassi, E. Capitani, M. Sciacco, A. Mangoni, and G. Scarlato.** 1992. Guillain-Barré syndrome associated with high titers of anti-GM1 antibodies. J. Neurol. Sci. **109:**200-206. doi: http://dx.doi.org/10.1016/0022-510X(92)90169-L.

29. **Tam, C. C., L. C. Rodrigues, I. Petersen, A. Islam, A. Hayward, S. J. O'Brien, and S. J. O'Brien.** 2006. Incidence of Guillain-Barré syndrome among patients with *Campylobacter* infection: a general practice research database study. J Infect Dis. **194:**95-97. http://dx.doi.org/10.1086/504294.

30. **Guillain, G., J. Barré, and A. Strohl.** 1916. Sur un syndrome de radiculo-nevrite avec hyperalbuminose du liquide cephalorachidien sans reaction cellulaire. Remarques sur les characteres clinique et graphique des reflexes tendinaux. . Bull Mem Soc Med Hop Paris. **40:**1462-1470.

31. **Ropper, A. H.** 1992. The Guillain–Barré syndrome. N. Engl. J. Med. **326:**1130-1136. doi: 10.1056/NEJM199204233261706. http://dx.doi.org/10.1056/NEJM199204233261706.

32. **Rhodes, K. M., and A. E. Tattersfield.** 1982. Guillain-Barré syndrome associated with *Campylobacter* infection. Br Med J (Clin Res Ed). **285:**173-174. doi: 10.1136/bmj.285.6336.173.

33. **Yuki, N., T. Taki, F. Inagaki, T. Kasama, M. Takahashi, K. Saito, S. Handa, and T. Miyatake.** 1993. A bacterium lipopolysaccharide that elicits Guillain-Barré syndrome has a GM1 ganglioside-like structure. J. Exp. Med. **178:**1771-1775. doi: 10.1084/jem.178.5.1771.

34. **Aspinall, G. O., A. G. McDonald, H. Pang, L. A. Kurjanczyk, and J. L. Penner.** 1994. Lipopolysaccharides of *Campylobacter jejuni* serotype O:19: Structures of core oligosaccharide regions from the serostrain and two bacterial isolates from patients with the Guillain-Barre syndrome. Biochemistry (N. Y.). **33:**241-249. doi: 10.1021/bi00167a032. http://dx.doi.org/10.1021/bi00167a032.

35. **Prendergast, M. M., A. J. Lastovica, and A. P. Moran.** 1998. Lipopolysaccharides from *Campylobacter jejuni* O:41 strains associated with Guillain-Barré ayndrome exhibit mimicry of GM1 ganglioside. Infect Immun. **66:**3649-3655.

36. **Yuki, N., M. Yamada, M. Koga, M. Odaka, K. Susuki, Y. Tagawa, S. Ueda, T. Kasama, A. Ohnishi, S. Hayashi, H. Takahashi, M. Kamijo, and K. Hirata.** 2001. Animal model of axonal Guillain-Barré syndrome induced by sensitization with GM1 ganglioside. Ann. Neurol. **49:**712-720. doi: 10.1002/ana.1012.

37. **Goodfellow, J. A., T. Bowes, K. Sheikh, M. Odaka, S. K. Halstead, P. D. Humphreys, E. R. Wagner, N. Yuki, K. Furukawa, K. Furukawa, J. J. Plomp, and H. J. Willison.** 2005. Overexpression of GD1a ganglioside sensitizes motor nerve terminals to anti-GD1a antibody-mediated injury in a model of acute motor axonal neuropathy. J. Neurosci. **25:**1620-1628. doi: 10.1523/JNEUROSCI.4279-04.2005.

38. **Yuki, N., K. Susuki, M. Koga, Y. Nishimoto, M. Odaka, K. Hirata, K. Taguchi, T. Miyatake, K. Furukawa, T. Kobata, and M. Yamada.** 2004. Carbohydrate mimicry between human ganglioside GM1 and *Campylobacter jejuni* lipooligosaccharide causes Guillain–Barré syndrome. Proceedings of the National Academy of Sciences of the United States of America. **101:**11404-11409. doi: 10.1073/pnas.0402391101.

39. **Yu, R. K., Y. Tsai, T. Ariga, and M. Yanagisawa.** 2011. Structures, biosynthesis, and functions of gangliosides-An overview. J Oleo Sci. **60:**537-544. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3684167/.

40. **Wim Ang, C., B. C. Jacobs, and J. D. Laman.** 2004. The Guillain–Barré syndrome: a true case of molecular mimicry. Trends Immunol. **25:**61-66. doi: http://dx.doi.org/10.1016/j.it.2003.12.004.

41. **Ang, C., J. Laman, H. Willison, E. Wagner, H. Endtz, M. De Klerk, A. Tio-Gillen, Van den Braak N, B. Jacobs, and P. Van Doorn.** 2002. Structure of *Campylobacter jejuni* lipopolysaccharides determines antiganglioside specificity and clinical features of Guillain-Barré, and Miller Fisher patients. J Peripher Nerv Syst. **7:**206-206. doi: 10.1046/j.1529-8027.2002.02026_4.x.

42. **Neisser, A., B. Schwerer, H. Bernheimer, and A. P. Moran.** 2000. Ganglioside-induced antiganglioside antibodies from a neuropathy patient cross-react with lipopolysaccharides of *Campylobacter jejuni* associated with Guillain–Barré syndrome. J Neuroimmunol. **102:**85-88.

43. **Ang, C. W., N. Yuki, B. C. Jacobs, M. Koga, P. A. Van Doorn, P. I. M. Schmitz, and F. G. A. Van Der Meché.** 1999. Rapidly progressive, predominantly motor Guillain-Barré syndrome with anti-GalNAc-GD1a antibodies. Neurology. **53:**2122-2122. doi: 10.1212/WNL.53.9.2122.

44. **Yuki, N., C. Wim Ang, M. Koga, B. C. Jacobs, P. A. Van Doorn, K. Hirata, and F. G. A. Van Der Meché.** 2000. Clinical features and response to treatment in Guillain-Barré syndrome associated with antibodies to GM1b ganglioside. Ann. Neurol. **47:**314-321. doi: 10.1002/1531-8249(200003)47:33.0.CO;2-C.

45. **Ogawara, K., S. Kuwabara, M. Mori, T. Hattori, M. Koga, and N. Yuki.** 2000. Axonal Guillain-Barré syndrome: Relation to anti-ganglioside antibodies and *Campylobacter jejuni* infection in Japan. Ann Neurol. **48:**624-631. doi: 10.1002/1531-8249(200010)48:43.0.CO;2-O.

46. **Koga, M., M. Gilbert, M. Takahashi, J. Li, S. Koike, K. Hirata, and N. Yuki.** 2006. Comprehensive analysis of bacterial risk factors for the development of Guillain-Barré Syndrome after *Campylobacter jejuni* enteritis. J Infect Dis. **193:**547-555. doi: 10.1086/499969.

47. **Euzeby, J.** List of prokaryotic names with standing in nomenclature. Genus *Campylobacter*. . doi: http://www.bacterio.net/campylobacter.html.

48. **Olkkola, S.** 2016. Antimicrobial resistance and its mechanisms among *Campylobacter coli* and *Campylobacter upsaliensis* with a special focus on streptomycin. Helsingin yliopisto.

49. **Gillespie, I., S. O'Brien, J. A. Frost, G. K. Adak, H. P. Le, A. V. Swan, M. J. Painter, and K. R. Neal.** 2002. A case-case comparison of *Campylobacter coli* and *Campylobacter jejuni* infection: A tool for generating hypotheses. Emerg Infect Dis. **8:**937. doi: 10.3201/eid0809.010187. http://wwwnc.cdc.gov/eid/article/8/9/01-0187.

50. **Doorduyn, Y.** 2010. Risk factors for indigenous *Campylobacter jejuni* and *Campylobacter coli* infections in The Netherlands: a case-control study. Epidemiol. Infect. **138:**1391.

51. **Gürtler, M., T. Alter, S. Kasimir, and K. Fehlhaber.** 2005. The importance of *Campylobacter coli* in human campylobacteriosis: prevalence and genetic characterization. Epidemiology & Infection. **133:**1081. doi: 10.1017/S0950268805004164. http://dx.doi.org/10.1017/S0950268805004164.

52. **Bassal, R., L. Lerner, L. Valinsky, V. Agmon, N. Peled, C. Block, N. Keller, Y. Keness, D. Taran, B. Shainberg, S. Ken-Dror, O. Treygerman, T. Rouach, S. Lowenthal, T. Shohat, and D. Cohen.** 2016. Trends in the epidemiology of

campylobacteriosis in Israel (1999–2012)
. Foodborne Pathog Dis. **13:**448-455.

53. **Nohra, A., A. Grinberg, A. C. Midwinter, J. C. Marshall, J. M. Collins-Emerson, and N. P. French.** 2016. Molecular epidemiology of *Campylobacter coli* strains isolated from different sources in New Zealand between 2005 and 2014. Appl Environ Microbiol. **82:**4363-4370. doi: 10.1128/AEM.00934-16.

54. **Tam, C. C., S. J. O'Brien, G. K. Adak, S. M. Meakins, and J. A. Frost.** 2003. *Campylobacter coli*—an important foodborne pathogen. J Infect. **47:**28-32. doi: https://doi.org/10.1016/S0163-4453(03)00042-2.

55. **Bessede, E., P. Lehours, L. Labadi, S. Bakiri, and F. Megraud.** 2014. Comparison of characteristics of patients infected by *Campylobacter jejuni, Campylobacter coli,* and *Campylobacter fetus*. J. Clin. Microbiol. **52:**328-330. doi: 10.1128/JCM.03029-13 [doi].

56. **Kärenlampi, R., H. Rautelin, D. Schönberg-Norio, L. Paulin, and M. Hänninen.** 2007. Longitudinal study of finnish *Campylobacter jejuni and C. coli* isolates from humans, using multilocus sequence typing, including comparison with epidemiological data and isolates from poultry and cattle. Applied and Environmental Microbiology. **73:**148-155. doi: 10.1128/AEM.01488-06.

57. **Roux, F., E. Sproston, O. Rotariu, M. MacRae, S. K. Sheppard, P. Bessell, A. Smith-Palmer, J. Cowden, M. C. J. Maiden, K. J. Forbes, and N. J. C. Strachan.** 2013. Elucidating the aetiology of human *Campylobacter coli* infections. PLoS ONE. **8:**e64504. http://dx.doi.org/10.1371%2Fjournal.pone.0064504.

58. **Horrocks, S. M., R. C. Anderson, D. J. Nisbet, and S. C. Ricke.** 2009. Incidence and ecology of *Campylobacter jejuni* and *coli* in animals. Anaerobe. **15:**18-25. doi: https://doi.org/10.1016/j.anaerobe.2008.09.001.

59. **Dingle, K. E., F. M. Colles, D. Falush, and M. C. Maiden.** 2005. Sequence typing and comparison of population biology of *Campylobacter coli* and *Campylobacter jejuni*. J. Clin. Microbiol. **43:**340-347. http://search.proquest.com/docview/288998084/abstract/embedded/MG047FZ8LPL0F5HQ?source=fedsrch.

60. **Zautner, A. E., A. Goldschmidt, A. Thürmer, J. Schuldes, O. Bader, R. Lugert, U. Groß, K. Stingl, G. Salinas, and T. Lingner.** 2015. SMRT sequencing of the *Campylobacter coli* BfR-CA-9557 genome sequence reveals unique methylation motifs. BMC Genomics. **16:**1088.

61. **Vorwerk, H., C. Huber, J. Mohr, B. Bunk, S. Bhuju, O. Wensel, C. Spröer, A. Fruth, A. Flieger, and K. Schmidt- Hohagen.** 2015. A transferable plasticity region in *Campylobacter coli* allows isolates of an otherwise non- glycolytic food- borne pathogen to catabolize glucose. Mol Microbiol. **98:**809-830.

62. **Méric, G., K. Yahara, L. Mageiros, B. Pascoe, M. C. Maiden, K. A. Jolley, and S. K. Sheppard.** 2014. A reference pan-genome approach to comparative bacterial

genomics: identification of novel epidemiological markers in pathogenic *Campylobacter*. PLoS ONE. **9:**e92798.

63. **Sheppard, S. K., N. D. McCarthy, D. Falush, and M. C. J. Maiden.** 2008. Convergence of *Campylobacter* Species: Implications for Bacterial Evolution. Science. **320:**237-239. doi: 10.1126/science.1155532.

64. **Sheppard, S. K.** 2010. Evolution of an agriculture-associated disease causing Campylobacter coli clade: evidence from national surveillance data in Scotland. PloS One. **5:**15708. http://search.proquest.com/docview/821197460/abstract/embedded/XQ9JEWJRFDG0NPLD?source=fedsrch.

65. **Thépault, A., G. Méric, K. Rivoal, B. Pascoe, L. Mageiros, F. Touzain, V. Rose, V. Béven, M. Chemaly, and S. K. Sheppard.** 2017. Genome-wide identification of host-segregating epidemiological markers for source attribution in *Campylobacter jejuni*. Appl Environ Microbiol. **83:**. doi: 10.1128/AEM.03085-16.

66. **Sheppard, S. K., X. Didelot, K. A. Jolley, A. E. Darling, B. Pascoe, G. Meric, D. J. Kelly, A. Cody, F. M. Colles, N. J. C. Strachan, I. D. Ogden, K. Forbes, N. P. French, P. Carter, W. G. Miller, N. D. McCarthy, R. Owen, E. Litrup, M. Egholm, J. P. Affourtit, S. D. Bentley, J. Parkhill, M. C. J. Maiden, and D. Falush.** 2013. Progressive genome-wide introgression in agricultural *Campylobacter coli*. Mol. Ecol. **22:**1051-1064. doi: 10.1111/mec.12162.

67. **Sheppard, S. K., and M. C. J. Maiden.** 2015. The evolution of *Campylobacter jejuni* and *Campylobacter coli*. Cold Spring Harb Perspect Biol. **7:**. doi: 10.1101/cshperspect.a018119.

68. **Dingle, K., F. Colles, D. Wareing, R. Ure, A. Fox, F. Bolton, H. Bootsma, R. Willems, R. Urwin, and M. Maiden.** 2001. Multilocus sequence typing system for *Campylobacter jejuni*. J. Clin. Microbiol. **39:**14-23. http://search.proquest.com/docview/17853837/abstract/embedded/19SDZKVUBDEOBEOZ?source=fedsrch.

69. **Miller, W. G., S. L. On, G. Wang, S. Fontanoz, A. J. Lastovica, and R. E. Mandrell.** 2005. Extended multilocus sequence typing system for *Campylobacter coli, C. lari, C. upsaliensis*, and *C. helveticus*. J. Clin. Microbiol. **43:**2315-2329. http://search.proquest.com/docview/289006504/abstract/embedded/N0P0NFBG5WZRIYNZ?source=fedsrch.

70. **Maiden, M., J. Bygraves, E. Feil, G. Morelli, J. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. Caugant, I. Feavers, M. Achtman, and B. Spratt.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U. S. A. **95:**3140-3145. http://search.proquest.com/docview/79735387/abstract/embedded/PT3U04NWLJ9XQVFH?source=fedsrch.

71. **Colles, F., and M. Maiden.** 2012. *Campylobacter* sequence typing databases: applications and future prospects. Microbiology. **158:**2695-2709.

72. **Hald, B., H. Skovgard, K. Pedersen, and H. Bunkenborg.** 2008. Influxed insects as vectors for *Campylobacter jejuni* and *Campylobacter coli* in Danish broiler houses. Poult Sci. **87:**1428-1434. doi: 10.3382/ps.2007-00301 [doi].

73. **Broman, T., S. Bergström, S. L. W. On, H. Palmgren, D. J. McCafferty, M. Sellin, and B. Olsen.** 2000. Isolation and characterization of *Campylobacter jejuni* subsp. *jejuni* from macaroni penguins (*Eudyptes chrysolophus*) in the subantarctic region. Appl Environ Microbiol. **66:**449-452. doi: 10.1128/AEM.66.1.449-452.2000.

74. **Griekspoor, P., F. M. Colles, N. D. McCarthy, P. M. Hansbro, C. Ashhurst-Smith, B. Olsen, D. Hasselquist, M. C. Maiden, and J. Waldenström.** 2013. Marked host specificity and lack of phylogeographic population structure of *Campylobacter jejuni* in wild birds. Mol Ecol. **22:**1463-1472.

75. **Sheppard, S., F. Colles, N. McCarthy, N. Strachan, L. Ogden, K. Forbes, J. Dallas, and C. Maiden.** 2011. Niche segregation and genetic structure of *Campylobacter jejuni* populations from wild and agricultural host species. Mol Ecol. **20:**3484-3490.

76. **Misawa, N., S. Shinohara, H. Satoh, H. Itoh, K. Shinohara, K. Shimomura, F. Kondo, and K. Itoh.** 2000. Isolation of *Campylobacter* species from zoo animals and polymerase chain reaction-based random amplified polymorphism DNA analysis. Vet Microbiol. **71:**59-68. doi: https://doi.org/10.1016/S0378-1135(99)00156-X.

77. **Colles, F. M., K. E. Dingle, A. J. Cody, and M. C. J. Maiden.** 2008. Comparison of *Campylobacter* populations in wild geese with those in starlings and free-range poultry on the same farm. Appl Environ Microbiol. **74:**3583-3590. doi: 10.1128/AEM.02491-07.

78. **Sheppard, S. K., F. Colles, J. Richardson, A. J. Cody, R. Elson, A. Lawson, G. Brick, R. Meldrum, C. L. Little, R. J. Owen, M. C. J. Maiden, and N. D. McCarthy.** 2010. Host association of *Campylobacter* genotypes transcends geographic variation. Applied and Environmental Microbiology. **76:**5269-5277. doi: 10.1128/AEM.00124-10.

79. **Sheppard, S. K., L. Cheng, G. Méric, C. Haan, A. Llarena, P. Marttinen, A. Vidal, A. Ridley, F. Clifton- Hadley, and T. R. Connor.** 2014. Cryptic ecology among host generalist *Campylobacter jejuni* in domestic animals. Mol Ecol. **23:**2442-2451.

80. **Gripp, E., D. Hlahla, X. Didelot, F. Kops, S. Maurischat, K. Tedin, T. Alter, and L. Ellerbroek.** 2011. Closely related *Campylobacter jejuni* strains from different sources reveal a generalist rather than a specialist lifestyle. BMC Genomics. **12:**584.

81. **McCarthy, N. D., I. A. Gillespie, A. J. Lawson, J. Richardson, K. R. Neal, P. R. Hawtin, M. C. J. Maiden, and S. J. O'Brien.** 2012. Molecular epidemiology of human *Campylobacter jejuni* shows association between seasonal and international patterns of disease. Epidemiol Infect. **140:**2247-2255. doi: 10.1017/S0950268812000192.

82. **Dearlove, B. L., A. J. Cody, B. Pascoe, G. Méric, D. J. Wilson, and S. K. Sheppard.** 2016. Rapid host switching in generalist *Campylobacter* strains erodes the signal for tracing human infections. The ISME Journal. **10:**721-729.

83. **de Haan, C. P. A., R. Kivistö, M. Hakkinen, H. Rautelin, and M. L. Hänninen.** 2010. Decreasing trend of overlapping multilocus sequence types between human and chicken *Campylobacter jejuni* isolates over a decade in Finland. Appl Environ Microbiol. **76:**5228-5236. doi: 10.1128/AEM.00581-10.

84. **Wimalarathna, H. M., and S. Sheppard.** 2012. Ecological factors that influence genetic structure in *Campylobacter coli* and *Campylobacter jejuni* , p. 1-8. *In* K. Anamthawat-Jónsson (ed.), Current topics in phylogenetics and phylogeography of terrestrial and aquatic systems, 1st ed., . InTech.

85. **Miller, W., M. Englen, S. Kathariou, I. Wesley, G. Wang, L. Pittenger-Alley, R. Siletz, W. Muraoka, P. Fedorka-Cray, and R. Mandrell.** 2006. **Identification of host-associated alleles by multilocus sequence typing of *Campylobacter coli* strains from food animals.** Microbiology. **152:**245-255.

86. **van Vliet, A. H., B. M. Pearson, N. J. Williams, B. Pascoe, G. Meric, P. Ashton, C. Jenkins, S. K. Sheppard, L. C. Crossman, and J. Wain.** 2016. Generating tools for the molecular epidemiology of *Campylobacter coli* by next generation genome sequencing. .

87. **Sheppard, S. K., N. D. McCarthy, K. A. Jolley, and M. C. Maiden.** 2011. Introgression in the genus *Campylobacter*: generation and spread of mosaic alleles. Microbiology. **157:**1066-1074.

88. **Muraoka, W. T., and Q. Zhang.** 2011. Phenotypic and genotypic evidence for L-fucose utilization by *Campylobacter jejuni*. J. Bacteriol. **193:**1065-1075. http://search.proquest.com/docview/851476283/abstract/embedded/L0Y0Y2HDMM23BEJS?source=fedsrch.

89. **Sopwith, W., A. Birtles, M. Matthews, A. Fox, S. Gee, S. James, J. Kempster, M. Painter, V. Edwards-Jones, K. Osborn, M. Regan, Q. Syed, and E. Bolton.** 2010. Investigation of food and environmental exposures relating to the epidemiology of *Campylobacter coli* in humans in Northwest England. Appl Environ Microbiol. **76:**129-135. doi: 10.1128/AEM.00942-09.

90. **Sheppard, S. K., J. F. Dallas, N. J. Strachan, M. MacRae, N. D. McCarthy, D. J. Wilson, F. J. Gormley, D. Falush, I. D. Ogden, M. C. Maiden, and K. J. Forbes.** 2009. *Campylobacter* genotyping to determine the source of human infection. Clin Infect Dis. **48:**1072-1078. doi: 10.1086/597402 [doi].

91. **Kittl, S., G. Heckel, B. M. Korczak, and P. Kuhnert.** 2013. Source attribution of human *Campylobacter* isolates by MLST and fla-typing and association of genotypes with quinolone resistance. PLoS ONE. **8:**e81796.

92. **Spaink, H. P., and D. M. Sheeley.** 1991. A novel highly unsaturated fatty acid moiety of lipo-oligosaccharide signals determines host specificity of *Rhizobium*. Nature. **354:**125.

93. **Michel, A., O. Clermont, E. Denamur, and O. Tenaillon.** 2010. Bacteriophage PhiX174's ecological niche and the flexibility of its *Escherichia coli* lipopolysaccharide receptor. Appl Environ Microbiol. **76:**7310-7313. doi: 10.1128/AEM.02721-09.

94. **Hynes, S. O., J. A. Ferris, B. Szponar, T. Wadström, J. G. Fox, J. O'Rourke, L. Larsson, E. Yaquian, Å Ljungh, M. Clyne, L. P. Andersen, and A. P. Moran.** 2004. Comparative chemical and biological characterization of the lipopolysaccharides of gastric and enterohepatic *Helicobacters*. Helicobacter. **9:**313-323. doi: 10.1111/j.1083-4389.2004.00237.x.

95. **Nilsson, C., A. Skoglund, A. P. Moran, H. Annuk, L. Engstrand, and S. Normark.** 2006. An enzymatic ruler modulates Lewis antigen glycosylation of *Helicobacter pylori* LPS during persistent infection. Proc Natl Acad Sci USA. **103:**2863-2868. doi: 10.1073/pnas.0511119103.

96. **Gilbert, M., J. Brisson, M. Karwaski, J. Michniewicz, A. Cunningham, Y. Wu, N. M. Young, and W. W. Wakarchuk.** 2000. Biosynthesis of ganglioside mimics in *Campylobacter jejuni* OH4384: identification of the glycosyltransferase genes, enzymatic synthesis of model compounds, and characterization of nanomole amounts by 600-MHz 1H and 13C NMR analysis . J Biol Chem. **275:**3896-3906. doi: 10.1074/jbc.275.6.3896.

97. **Semchenko, E. A., C. J. Day, M. Moutin, J. C. Wilson, J. Tiralongo, and V. Korolik.** 2012. Structural heterogeneity of terminal glycans in *Campylobacter jejuni* lipooligosaccharides. Plos One. **7:**e40920. http://dx.doi.org/10.1371%2Fjournal.pone.0040920.

98. **Houliston, R., S. Bernatchez, M. Karwaski, R. E. Mandrell, H. C. Jarrell, W. W. Wakarchuk, and M. Gilbert.** 2009. Complete chemoenzymatic synthesis of the Forssman antigen using novel glycosyltransferases identified in *Campylobacter jejuni* and *Pasteurella multocida*. Glycobiology. **19:**153-159. http://dx.doi.org/10.1093/glycob/cwn117.

99. **Mathé, C., M. Sagot, T. Schiex, and P. Rouzé.** 2002. Current methods of gene prediction, their strengths and weaknesses. Nucleic Acids Res. **30:**4103-4117. http://dx.doi.org/10.1093/nar/gkf543.

100. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**. doi: 10.1093/nar/25.17.3389. http://dx.doi.org/10.1093/nar/25.17.3389.

101. **Koski, L. B., and G. B. Golding.** 2001. The Closest BLAST Hit Is Often Not the Nearest Neighbor. J Mol Evol. **52:**540-542. doi: 10.1007/s002390010184. http://dx.doi.org/10.1007/s002390010184.

102. **Fitch, W. M.** 1970. Distinguishing homologous from analogous proteins. Syst. Biol. **19:**99-113. http://dx.doi.org/10.2307/2412448.

103. **Enright, A. J., S. Van Dongen, and C. A. Ouzounis.** 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Research. **30:**1575-1584. doi: 10.1093/nar/30.7.1575.

104. **Ekseth, O. K., M. Kuiper, and V. Mironov.** 2013. orthAgogue: an agile tool for the rapid prediction of orthology relations. Bioinformatics. . doi: 10.1093/bioinformatics/btt582.

105. **Page, A. J., C. A. Cummins, M. Hunt, V. K. Wong, S. Reuter, M. T. G. Holden, M. Fookes, D. Falush, J. A. Keane, and J. Parkhill.** 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. **31:**3691-3693. http://dx.doi.org/10.1093/bioinformatics/btv421.

106. **Bernatchez, S., M. Gilbert, M. Blanchard, M. Karwaski, J. Li, S. DeFrees, and W. W. Wakarchuk.** 2007. Variants of the β1,3-galactosyltransferase CgtB from the bacterium *Campylobacter jejuni* have distinct acceptor specificities. Glycobiology. **17:**1333-1343. http://dx.doi.org/10.1093/glycob/cwm090.

107. **Bornscheuer, U. T., and R. J. Kazlauskas.** 2004. Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways. Angew Chem Int Ed. **43:**6032-6040.

108. **Gerlt, J. A., and P. C. Babbitt.** 2000. Can sequence determine function? Genome Biol. **1:**reviews0005.1. doi: 10.1186/gb-2000-1-5-reviews0005. http://dx.doi.org/10.1186/gb-2000-1-5-reviews0005.

109. **Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler, and R. Lopez.** 2005. InterProScan: protein domains identifier. Nucleic Acids Res. **33:**W116-W120. http://dx.doi.org/10.1093/nar/gki442.

110. **Laine, R. A.** 1994. Invited Commentary: A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05 × 1012 structures for a reducing hexasaccharide: the *Isomer Barrier* to development of single-method saccharide sequencing or synthesis systems. Glycobiology. **4:**759-767. http://dx.doi.org/10.1093/glycob/4.6.759.

111. **Campbell, J. A., G. J. Davies, V. Bulone, and B. Henrissat.** 1997. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem. J. **326:**929-939. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1218753/.

112. **Lairson, L. L., B. Henrissat, G. J. Davies, and S. G. Withers.** 2008. Glycosyltransferases: structures, functions, and mechanisms. Annu Rev Biochem. **77:**521-555. doi: 10.1146/annurev.biochem.76.061005.092322. http://dx.doi.org/10.1146/annurev.biochem.76.061005.092322.

113. **Bourne, Y., and B. Henrissat.** 2001. Glycoside hydrolases and glycosyltransferases: families and functional modules. Curr Opin Struct Biol. **11:**593-600. doi: http://doi.org/10.1016/S0959-440X(00)00253-0.

114. **Coutinho, P. M., E. Deleury, G. J. Davies, and B. Henrissat.** 2003. An evolving hierarchical family classification for glycosyltransferases. J Mol Biol. **328:**307-317. doi: http://doi.org/10.1016/S0022-2836(03)00307-3.

115. **Yin, Y., X. Mao, J. Yang, X. Chen, F. Mao, and Y. Xu.** 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. **40:**W445-W451. http://dx.doi.org/10.1093/nar/gks479.

116. **Park, B. H., T. V. Karpinets, M. H. Syed, M. R. Leuze, and E. C. Uberbacher.** 2010. CAZymes Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. Glycobiology. **20:**1574-1584. http://dx.doi.org/10.1093/glycob/cwq106.

117. **Busk, P. K., B. Pilgaard, M. J. Lezyk, A. S. Meyer, and L. Lange.** 2017. Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function. BMC Bioinformatics. **18:**214. doi: 10.1186/s12859-017-1625-9. http://dx.doi.org/10.1186/s12859-017-1625-9.

118. **Chan, P. H. W., L. L. Lairson, H. J. Lee, W. W. Wakarchuk, N. C. J. Strynadka, S. G. Withers, and L. P. McIntosh.** 2009. NMR spectroscopic characterization of the sialyltransferase CstII from *Campylobacter jejuni*: histidine 188 is the general base. Biochemistry (N. Y.). **48:**11220-11230. doi: 10.1021/bi901606n. http://dx.doi.org/10.1021/bi901606n.

119. **Chiu, C. P. C., A. G. Watts, L. L. Lairson, M. Gilbert, D. Lim, W. W. Wakarchuk, S. G. Withers, and N. C. J. Strynadka.** 2004. Structural analysis of the sialyltransferase CstII from *Campylobacter jejuni* in complex with a substrate analog. Nat Struct Mol Biol. **11:**163-170. http://dx.doi.org/10.1038/nsmb720.

120. **Cheng, J., H. Yu, K. Lau, S. Huang, H. A. Chokhawala, Y. Li, V. K. Tiwari, and X. Chen.** 2008. Multifunctionality of *Campylobacter jejuni* sialyltransferase CstII: characterization of GD3/GT3 oligosaccharide synthase, GD3 oligosaccharide sialidase, and trans-sialidase activities. Glycobiology. **18:**686-697. http://dx.doi.org/10.1093/glycob/cwn047.

121. **Glover, K. J., E. Weerapana, S. Numao, and B. Imperiali.** 2005. Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*. Chem Biol. **12:**1311-1316. doi: http://doi.org/10.1016/j.chembiol.2005.10.004.

122. **Linton, D., N. Dorrell, P. G. Hitchen, S. Amber, A. V. Karlyshev, H. R. Morris, A. Dell, M. A. Valvano, M. Aebi, and B. W. Wren.** 2005. Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. Mol Microbiol. **55:**1695-1703. doi: 10.1111/j.1365-2958.2005.04519.x.

123. **Glover, K. J., E. Weerapana, and B. Imperiali.** 2005. In vitro assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked glycosylation. Proc Natl Acad Sci U S A. **102:**14255-14259. doi: 10.1073/pnas.0507311102.

124. **Troutman, J. M., and B. Imperiali.** 2009. *Campylobacter jejuni* PglH is a single active site processive polymerase that utilizes product inhibition to limit sequential glycosyl transfer reactions. Biochemistry (N. Y.). **48:**2807-2816. doi: 10.1021/bi802284d. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2736683/.

125. **Chiu, C. P. C., L. L. Lairson, M. Gilbert, W. W. Wakarchuk, S. G. Withers, and N. C. J. Strynadka.** 2007. Structural analysis of the a-2,3-sialyltransferase Cst-I from *Campylobacter jejuni* in apo and substrate-analogue bound forms. Biochemistry. **46:**7196-7204. doi: 10.1021/bi602543d. http://dx.doi.org/10.1021/bi602543d.

126. **Ewing, C. P., E. Andreishcheva, and P. Guerry.** 2009. Functional characterization of flagellin glycosylation in *Campylobacter jejuni* 81-176. J Bacteriol. **191:**7086-7093. doi: 10.1128/JB.00378-09 [doi].

127. **Thibault, P., S. M. Logan, J. F. Kelly, J. Brisson, C. P. Ewing, and P. Guerry.** 2001. Identification of the carbohydrate moieties and glycosylation motifs in *Campylobacter jejuni* flagellin. J Biol Chem. **276:**34862-34870.

128. **Guerry, P., C. P. Ewing, M. Schirm, M. Lorenzo, J. Kelly, D. Pattarini, G. Majam, P. Thibault, and S. Logan.** 2006. Changes in flagellin glycosylation affect *Campylobacter* autoagglutination and virulence. Mol Microbiol. **60:**299-311.

129. **Goon, S., J. F. Kelly, S. M. Logan, C. P. Ewing, and P. Guerry.** 2003. Pseudaminic acid, the major modification on *Campylobacter* flagellin, is synthesized via the *Cj1293* gene. Mol Microbiol. **50:**659-671.

130. **Schirm, M., I. C. Schoenhofen, S. M. Logan, K. C. Waldron, and P. Thibault.** 2005. Identification of unusual bacterial glycosylation by tandem mass spectrometry analyses of intact proteins. Anal. Chem. **77:**7774-7782.

131. **Schoenhofen, I. C., D. J. McNally, J. Brisson, and S. M. Logan.** 2006. Elucidation of the CMP-pseudaminic acid pathway in *Helicobacter pylori*: synthesis from UDP-N-acetylglucosamine by a single enzymatic reaction. Glycobiology. **16:**8C-14C. doi: 10.1093/glycob/cwl010.

132. **Zebian, N., A. Merkx-Jacques, P. P. Pittock, S. Houle, C. M. Dozois, G. A. Lajoie, and C. Creuzenet.** 2015. Comprehensive analysis of flagellin glycosylation in *Campylobacter jejuni* NCTC 11168 reveals incorporation of legionaminic acid and its importance for host colonization. Glycobiology. **26:**386-397.

133. **Ulasi, G. N., A. J. Creese, S. X. Hui, C. W. Penn, and H. J. Cooper.** 2015. Comprehensive mapping of O- glycosylation in flagellin from *Campylobacter jejuni* 11168: A multienzyme differential ion mobility mass spectrometry approach. Proteomics. **15:**2733-2745.

134. **Logan, S. M., J. P. Hui, E. Vinogradov, A. J. Aubry, J. E. Melanson, J. F. Kelly, H. Nothaft, and E. C. Soo.** 2009. Identification of novel carbohydrate modifications on *Campylobacter jejuni* 11168 flagellin using metabolomics- based approaches. The FEBS Journal. **276:**1014-1023.

135. **Howard, S. L., A. Jagannathan, E. C. Soo, J. P. Hui, A. J. Aubry, I. Ahmed, A. Karlyshev, J. F. Kelly, M. A. Jones, M. P. Stevens, S. M. Logan, and B. W. Wren.** 2009. *Campylobacter jejuni* glycosylation island important in cell charge, legionaminic acid biosynthesis, and colonization of chickens. Infect Immun. **77:**2544-2556. doi: 10.1128/IAI.01425-08 [doi].

136. **McNally, D. J., A. J. Aubry, J. P. Hui, N. H. Khieu, D. Whitfield, C. P. Ewing, P. Guerry, J. R. Brisson, S. M. Logan, and E. C. Soo.** 2007. Targeted metabolomics analysis of *Campylobacter coli* VC167 reveals legionaminic acid derivatives as novel flagellar glycans. J Biol Chem. **282:**14463-14475. doi: M611027200 [pii].

137. **Logan, S. M., J. F. Kelly, P. Thibault, C. P. Ewing, and P. Guerry.** 2002. Structural heterogeneity of carbohydrate modifications affects serospecificity of *Campylobacter* flagellins. Mol Microbiol. **46:**587-597.

138. **Schoenhofen, I. C., E. Vinogradov, D. M. Whitfield, J. Brisson, and S. M. Logan.** 2009. The CMP-legionaminic acid pathway in *Campylobacter*: Biosynthesis involving novel GDP-linked precursors. Glycobiology. **19:**715-725. doi: 10.1093/glycob/cwp039.

139. **Young, N. M., J. Brisson, J. Kelly, D. C. Watson, L. Tessier, P. H. Lanthier, H. C. Jarrell, N. Cadotte, F. S. Michael, and E. Aberg.** 2002. Structure of the *N*-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, *Campylobacter jejuni*. J Biol Chem. **277:**42530-42539.

140. **Wacker, M., D. Linton, P. G. Hitchen, M. Nita-Lazar, S. M. Haslam, S. J. North, M. Panico, H. R. Morris, A. Dell, B. W. Wren, and M. Aebi.** 2002. *N*-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. Science. **298:**1790-1793. doi: 10.1126/science.298.5599.1790 [doi].

141. **Jervis, A. J., J. A. Butler, A. J. Lawson, R. Langdon, B. W. Wren, and D. Linton.** 2012. Characterization of the structurally diverse *N*-linked glycans of *Campylobacter* species. J Bacteriol. **194:**2355-2362. doi: 10.1128/JB.00042-12 [doi].

142. **Szymanski, C. M., R. Yao, C. P. Ewing, T. J. Trust, and P. Guerry.** 1999. Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. Mol Microbiol. **32:**1022-1030.

143. **Szymanski, C. M., F. S. Michael, H. C. Jarrell, J. Li, M. Gilbert, S. Larocque, E. Vinogradov, and J. Brisson.** 2003. Detection of conserved N-Linked glycans and phase-variable lipooligosaccharides and capsules from *Campylobacter* cells by mass spectrometry and high resolution magic angle spinning NMR spectroscopy. J Biol Chem. **278:**24509-24520. doi: 10.1074/jbc.M301273200.

144. **Szymanski, C. M., S. M. Logan, D. Linton, and B. W. Wren.** 2003. *Campylobacter*–a tale of two protein glycosylation systems. Trends Microbiol. **11:**233-238.

145. **Scott, N. E., N. B. Marzook, J. A. Cain, N. Solis, M. Thaysen-Andersen, S. P. Djordjevic, N. H. Packer, M. R. Larsen, and S. J. Cordwell.** 2014. Comparative

proteomics and glycoproteomics reveal increased *N*-linked glycosylation and relaxed sequon specificity in *Campylobacter jejuni* NCTC11168 O. J Proteome Res. **13:**5136-5150.

146. **Scott, N. E., B. L. Parker, A. M. Connolly, J. Paulech, A. V. Edwards, B. Crossett, L. Falconer, D. Kolarich, S. P. Djordjevic, P. Hojrup, N. H. Packer, M. R. Larsen, and S. J. Cordwell.** 2011. Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the *N*-linked glycoproteome of *Campylobacter jejuni*. Mol Cell Proteomics. **10:**M000031-MCP201. doi: 10.1074/mcp.M000031-MCP201 [doi].

147. **Alemka, A., H. Nothaft, J. Zheng, and C. M. Szymanski.** 2013. *N*-glycosylation of *Campylobacter jejuni* surface proteins promotes bacterial fitness. Infect Immun. **81:**1674-1682. doi: 10.1128/IAI.01370-12.

148. **Chart, H., J. Frost, A. Oza, R. Thwaites, S. Gillanders, and B. Rowe.** 1996. Heat- stable serotyping antigens expressed by strains of *Campylobacter jejuni* are probably capsular and not long- chain lipopolysaccharide. J Appl Microbiol. **81:**635-640.

149. **Karlyshev, A. V., D. Linton, N. A. Gregson, A. J. Lastovica, and B. W. Wren.** 2000. Genetic and biochemical evidence of a *Campylobacter jejuni* capsular polysaccharide that accounts for Penner serotype specificity. Mol Microbiol. **35:**529-541.

150. **Karlyshev, A. V., O. L. Champion, C. Churcher, J. Brisson, H. C. Jarrell, M. Gilbert, D. Brochu, F. St Michael, J. Li, W. W. Wakarchuk, I. Goodhead, M. Sanders, K. Stevens, B. White, J. Parkhill, B. W. Wren, and C. M. Szymanski.** 2005. Analysis of *Campylobacter jejuni* capsular loci reveals multiple mechanisms for the generation of structural diversity and the ability to form complex heptoses. Mol. Microbiol. **55:**90-103. doi: 10.1111/j.1365-2958.2004.04374.x.

151. **Bernatchez, S., C. M. Szymanski, N. Ishiyama, J. Li, H. C. Jarrell, P. C. Lau, A. M. Berghuis, N. M. Young, and W. W. Wakarchuk.** 2005. A single bifunctional UDP-GlcNAc/Glc 4-epimerase supports the synthesis of three cell surface glycoconjugates in *Campylobacter jejuni*. J. Biol. Chem. **280:**4792-4802. doi: 10.1074/jbc.M407767200.

152. **Naito, M., E. Frirdich, J. A. Fields, M. Pryjma, J. Li, A. Cameron, M. Gilbert, S. A. Thompson, and E. C. Gaynor.** 2010. Effects of sequential *Campylobacter jejuni* 81-176 lipooligosaccharide core truncations on biofilm formation, stress survival, and pathogenesis. Journal of Bacteriology. **192:**2182-2192.

153. **Reife, R. A., S. R. Coats, M. Al-Qutub, D. M. Dixon, P. A. Braham, R. J. Billharz, W. N. Howald, and R. P. Darveau.** 2006. *Porphyromonas gingivalis* lipopolysaccharide lipid A heterogeneity: differential activities of tetra- and penta-acylated lipid A structures on E-selectin expression and TLR4 recognition. Cell Microbiol. **8:**857-868. doi: 10.1111/j.1462-5822.2005.00672.x.

154. **Trent, M. S., C. M. Stead, A. X. Tran, and J. V. Hankins.** 2006. Diversity of endotoxin and its impact on pathogenesis. J Endotoxin Res. **12:**205-223. doi:

10.1177/09680519060120040201.
http://journals.sagepub.com/doi/abs/10.1177/09680519060120040201.

155. **Preston, A., R. E. Mandrell, B. W. Gibson, and M. A. Apicella.** 1996. The lipooligosaccharides of pathogenic Gram-negative bacteria. Crit Rev Microbiol. **22:**139-180.

156. **Rathinam, V. A., D. M. Appledorn, K. A. Hoag, A. Amalfitano, and L. S. Mansfield.** 2009. *Campylobacter jejuni*-induced activation of dendritic cells involves cooperative signaling through Toll-Like Receptor 4 (TLR4)-MyD88 and TLR4-TRIF axes. Infect. Immun. **77:**2499-2507. doi: 10.1128/IAI.01562-08.

157. **Jerala, R.** 2007. Structural biology of the LPS recognition. Int J Med Microbiol. **297:**353-363.

158. **Lu, Y., W. Yeh, and P. S. Ohashi.** 2008. LPS/TLR4 signal transduction pathway. Cytokine. **42:**145-151.

159. **Park, B. S., D. H. Song, H. M. Kim, B. Choi, H. Lee, and J. Lee.** 2009. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nature. **458:**1191-1195. http://dx.doi.org/10.1038/nature07830.

160. **Kuijf, M. L., J. N. Samsom, W. van Rijs, M. Bax, R. Huizinga, A. P. Heikema, P. A. van Doorn, A. van Belkum, Y. van Kooyk, P. C. Burgers, T. M. Luider, H. P. Endtz, E. E. S. Nieuwenhuis, and B. C. Jacobs.** 2010. TLR4-mediated sensing of *Campylobacter jejuni* by dendritic cells is determined by sialylation. J. Immunol. **185:**748-755. doi: 10.4049/jimmunol.0903014.

161. **Bax, M., M. L. Kuijf, A. P. Heikema, W. van Rijs, S. C. M. Bruijns, J. J. García-Vallejo, P. R. Crocker, B. C. Jacobs, S. J. van Vliet, and Y. van Kooyk.** 2011. *Campylobacter jejuni* lipooligosaccharides modulate dendritic cell-mediated T cell polarization in a sialic acid linkage-dependent manner. Infection and Immunity. **79:**2681-2689. doi: 10.1128/IAI.00009-11.

162. **Huizinga, R., W. van Rijs, J. J. Bajramovic, M. L. Kuijf, J. D. Laman, J. N. Samsom, and B. C. Jacobs.** 2013. Sialylation of *Campylobacter jejuni* endotoxin promotes dendritic cell–mediated B cell responses through CD14-dependent production of IFN-ß and TNF-a. J. Immunol. **191:**5636-5645. doi: 10.4049/jimmunol.1301536.

163. **Huizinga, R., A. S. Easton, A. M. Donachie, J. Guthrie, W. van Rijs, A. Heikema, L. Boon, J. N. Samsom, B. C. Jacobs, H. J. Willison, and C. S. Goodyear.** 2012. Sialylation of *Campylobacter jejuni* lipo-oligosaccharides: Impact on phagocytosis and cytokine production in mice. PLoS ONE. **7:**e34416. http://dx.doi.org/10.1371%2Fjournal.pone.0034416.

164. **Stephenson, H. N., C. M. John, N. Naz, O. Gundogdu, N. Dorrell, B. W. Wren, G. A. Jarvis, and M. Bajaj-Elliott.** 2013. *Campylobacter jejuni* lipooligosaccharide sialylation, phosphorylation, and amide/ester linkage modifications fine-tune human toll-

like receptor 4 activation. J. Biol. Chem. **288:**19661-19672. doi: 10.1074/jbc.M113.468298.

165. **Taboada, E. N., A. van Belkum, N. Yuki, R. R. Acedillo, P. C. R. Godschalk, M. Koga, H. P. Endtz, M. Gilbert, and J. H. E. Nash.** 2007. Comparative genomic analysis of *Campylobacter jejuni* associated with Guillain-Barré and Miller Fisher syndromes: neuropathogenic and enteritis-associated isolates can share high levels of genomic similarity. BMC Genomics. **8:**359. doi: 10.1186/1471-2164-8-359. http://dx.doi.org/10.1186/1471-2164-8-359.

166. **Louwen, R., D. Horst-kreft, A. de Boer, L. van der Graaf, G. de Knegt, M. Hamersma, A. Heikema, A. Timms, B. Jacobs, J. Wagenaar, H. Endtz, J. van der Oost, J. Wells, E. Nieuwenhuis, A. van Vliet, P. Willemsen, P. van Baarlen, and A. van Belkum.** 2013. A novel link between *Campylobacter jejuni* bacteriophage defence, virulence and Guillain-Barré syndrome. European Journal of Clinical Microbiology and Infectious Diseases. **32:**207-226. http://search.proquest.com/docview/1271881022/abstract/embedded/EAKIJ5EBLDEZ1FZD?source=fedsrch.

167. **Islam, Z., A. van Belkum, J. A. Wagenaar, A. J. Cody, A. G. de Boer, H. Tabor, B. C. Jacobs, K. A. Talukder, and H. P. Endtz.** 2009. Comparative genotyping of *Campylobacter jejuni* strains from patients with Guillain-Barré syndrome in Bangladesh. Plos One. **4:**e7257. http://dx.doi.org/10.1371%2Fjournal.pone.0007257.

168. **Linton, D., A. V. Karlyshev, and B. W. Wren.** 2001. Deciphering *Campylobacter jejuni* cell surface interactions from the genome sequence. Curr Opin Microbiol. **4:**35-40. doi: http://dx.doi.org/10.1016/S1369-5274(00)00161-2.

169. **Linton, D., M. Gilbert, P. G. Hitchen, A. Dell, H. R. Morris, W. W. Wakarchuk, N. A. Gregson, and B. W. Wren.** 2000. Phase variation of a β-1,3 galactosyltransferase involved in generation of the ganglioside GM1-like lipo-oligosaccharide of *Campylobacter jejuni*. Mol. Microbiol. **37:**501-514. doi: 10.1046/j.1365-2958.2000.02020.x.

170. **Dorrell, N., J. A. Mangan, K. G. Laing, J. Hinds, D. Linton, H. Al-Ghusein, B. G. Barrell, J. Parkhill, N. G. Stoker, A. V. Karlyshev, P. D. Butcher, and B. W. Wren.** 2001. Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. Genome Res. **11:**1706-1715. doi: 10.1101/gr.185801.

171. **Pearson, B. M., C. Pin, J. Wright, K. I'Anson, T. Humphrey, and J. M. Wells.** 2003. Comparative genome analysis of *Campylobacter jejuni* using whole genome DNA microarrays. FEBS Lett. **554:**224-230. doi: 10.1016/S0014-5793(03)01164-5.

172. **Taboada, E. N., R. R. Acedillo, C. D. Carrillo, W. A. Findlay, D. T. Medeiros, O. L. Mykytczuk, M. J. Roberts, C. A. Valencia, J. M. Farber, and J. H. E. Nash.** 2004. Large-scale comparative genomics meta-analysis of *Campylobacter jejuni* isolates reveals low level of genome plasticity. J Clin Microbiol. **42:**4566-4576. doi: 10.1128/JCM.42.10.4566-4576.2004.

173. **Lang, P., T. Lefebure, W. Wang, P. Pavinski Bitar, R. J. Meinersmann, K. Kaya, and M. J. Stanhope.** 2010. Expanded multilocus sequence typing and comparative genomic hybridization of *Campylobacter coli* isolates from multiple hosts. Applied and Environmental Microbiology. **76:**1913. http://aem.asm.org/cgi/content/abstract/76/6/1913; http://aem.asm.org/cgi/reprint/76/6/1913.pdf.

174. **Parker, C. T., S. T. Horn, M. Gilbert, W. G. Miller, D. L. Woodward, and R. E. Mandrell.** 2005. Comparison of *Campylobacter jejuni* lipooligosaccharide biosynthesis loci from a variety of sources. J. Clin. Microbiol. **43:**2771-2781. doi: 10.1128/JCM.43.6.2771-2781.2005.

175. **Parker, C. T., M. Gilbert, N. Yuki, H. P. Endtz, and R. E. Mandrell.** 2008. Characterization of lipooligosaccharide-biosynthetic loci of *Campylobacter jejuni* reveals new lipooligosaccharide classes: evidence of mosaic organizations , J. Bacteriol. **190:**5681-5689.

176. **Phongsisay, V., V. N. Perera, and B. N. Fry.** 2006. Exchange of lipooligosaccharide synthesis genes creates potential Guillain-Barré syndrome-inducible strains of *Campylobacter jejuni*. Infect Immun. **74:**1368-1372. doi: 10.1128/IAI.74.2.1368-1372.2006.

177. **de Haan, C. P. A., A. Llarena, J. Revez, and M. Hänninen.** 2012. Association of *Campylobacter jejuni* Metabolic Traits with Multilocus Sequence Types. Applied and Environmental Microbiology. **78:**5550-5554. doi: 10.1128/AEM.01023-12.

178. **Revez, J., M. Rossi, P. Ellström, C. de Haan, H. Rautelin, and M. Hänninen.** 2011. Finnish *Campylobacter jejuni s*trains of multilocus sequence type ST-22 complex have two lineages with different characteristics. PLoS ONE. **6:**e26880.

179. **Revez, J., and M. Hänninen.** 2012. Lipooligosaccharide locus classes are associated with certain *Campylobacter jejuni* multilocus sequence types. Eur J Clin Microbiol Infect Dis. **31:**2203-2209.

180. **Habib, I.** 2009. Correlation between genotypic diversity, lipooligosaccharide gene locus class variation, and caco-2 cell invasion potential of Campylobacter jejuni isolates from chicken meat and humans: contribution to virulotyping. Appl. Environ. Microbiol. **75:**4277.

181. **HOTTER, G. S., I. H. LI, and N. P. FRENCH.** 2010. Binary genomotyping using lipooligosaccharide biosynthesis genes distinguishes between Campylobacter jejuni isolates within poultry-associated multilocus sequence types. Epidemiol. Infect. **138:**992-1003. doi: http://dx.doi.org.libproxy.helsinki.fi/10.1017/S0950268809991075. http://search.proquest.com.libproxy.helsinki.fi/docview/347023296?accountid=11365.

182. **Bayliss, C. D.** 2009. Determinants of phase variation rate and the fitness implications of differing rates for bacterial pathogens and commensals. FEMS Microbiol. Rev. **33:**504-520. doi: 10.1111/j.1574-6976.2009.00162.x [doi].

183. **Hallet, B.** 2001. Playing Dr Jekyll and Mr Hyde: combined mechanisms of phase variation in bacteria. Curr. Opin. Microbiol. **4:**570-581.

184. **Moxon, R., C. Bayliss, and D. Hood.** 2006. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. Annu Rev Genet. **40:**307-333.

185. **Orsi, R. H., B. M. Bowen, and M. Wiedmann.** 2010. Homopolymeric tracts represent a general regulatory mechanism in prokaryotes. BMC Genomics. **11:**102.

186. **Gilbert, M., P. C. Godschalk, C. T. Parker, H. P. Endtz, and W. W. Wakarchuk.** 2005. Genetic bases for the variation in the lipooligosaccharide outer core of Campylobacter jejuni and possible association of glycosyltransferase genes with post-infectious neuropathies, p. 219-248. *In* J. Ketley and M. Konkel (eds.), *Campylobacter:* Molecular and cellular biology. Horizon Scientific Press, Norwich, UK.

187. **Park, S. F., D. Purdy, and S. Leach.** 2000. Localized reversible frameshift mutation in the *flh*A gene confers phase variability to flagellin gene expression in *Campylobacter coli*. Journal of Bacteriology. **182:**207-210.

188. **Michael, F. S., C. M. Szymanski, J. Li, K. H. Chan, N. H. Khieu, S. Larocque, W. W. Wakarchuk, J. Brisson, and M. A. Monteiro.** 2002. The structures of the lipooligosaccharide and capsule polysaccharide of Campylobacter jejuni genome sequenced strain NCTC 11168. European Journal of Biochemistry. **269:**5119-5136.

189. **Aspinall, G. O., A. G. McDonald, H. Pang, L. A. Kurjanczyk, and J. L. Penner.** 1993. Lipopolysaccharide of *Campylobacter coli* serotype O:30. Fractionation and structure of liberated core oligosaccharide. J. Biol. Chem. **268:**6263-6268.

190. **Ernst, R. K., C. Y. Eugene, L. Guo, K. B. Lim, J. L. Burns, M. Hackett, and S. I. Miller.** 1999. Specific lipopolysaccharide found in cystic fibrosis airway *Pseudomonas aeruginosa*. Science. **286:**1561-1565.

191. **Velasco, J., H. Moll, Y. A. Knirel, V. Sinnwell, I. Moriyón, and U. Zähringer.** 1998. Structural studies on the lipopolysaccharide from a rough strain of *Ochrobactrum anthropi* containing a 2,3-diamino-2,3-dideoxy-d-glucose disaccharide lipid A backbone. Carbohydr Res. **306:**283-290. doi: http://dx.doi.org/10.1016/S0008-6215(97)10029-5.

192. **Sweet, C. R., A. H. Williams, M. J. Karbarz, C. Werts, S. R. Kalb, R. J. Cotter, and C. R. Raetz.** 2004. Enzymatic synthesis of lipid a molecules with four amide-linked acyl chains LpxA acyltransferases selective for an analog of UDP-N-acetylglucosamine in which an amine replaces the 3 ″-hydroxyl group. J Biol Chem. **279:**25411-25419.

193. **Moran, A. P., U. Zähringer, U. Seydel, D. Scholz, P. Stütz, and E. T. Rietschel.** 1991. Structural analysis of the lipid A component of *Campylobacter jejuni* CCUG 10936 (serotype O:2) lipopolysaccharide. Eur. J. Biochem. **198:**459-469. doi: 10.1111/j.1432-1033.1991.tb16036.x.

194. **Moran, A. P., D. T. O'Malley, T. U. Kosunen, and I. M. Helander.** 1994. Biochemical characterization of *Campylobacter fetus* lipopolysaccharides. Infect. Immun. **62:**3922-3929.

195. **Moran, A. P.** 1997. Structure and conserved characteristics of *Campylobacter jejuni* lipopolysaccharides. J. Infect. Dis. **176:**S115-S121. doi: 10.1086/513781.

196. **Cullen, T. W., and M. S. Trent.** 2010. A link between the assembly of flagella and lipooligosaccharide of the Gram-negative bacterium *Campylobacter jejuni*. Proc Natl Acad Sci USA. **107:**5160-5165. doi: 10.1073/pnas.0913451107 [doi].

197. **Cullen, T. W., J. P. O'Brien, D. R. Hendrixson, D. K. Giles, R. I. Hobb, S. A. Thompson, J. S. Brodbelt, and M. S. Trent.** 2013. *EptC* of *Campylobacter jejuni* mediates phenotypes involved in host interactions and virulence. Infect. Immun. **81:**430-440.

198. **Imoto, M., S. Kusumoto, T. Shiba, E. T. Rietschel, C. Galanos, and O. Lüderitz.** 1985. Chemical structure of *Escherichia coli* lipid A. Tetrahedron Lett. **26:**907-908. doi: http://dx.doi.org/10.1016/S0040-4039(00)61961-5.

199. **Naess, V.** 1985. In vitro blastogenic response and immunoglobulin synthesis in murine spleen cells stimulated with lipopolysaccharides from *Campylobacter jejuni* and *Campylobacter coli*. Acta Path Microbiol Immunol Scand. **93:**211-216.

200. **Branquinho, M., C. Alviano, and I. Ricciardi.** 1983. Determinacao da composicao quimica e acao biologica do lipopolissacarideo (LPS) de *Campylobacter fetus* ss. *jejuni*. Rev Microbiol. **14:**90-96.

201. **Sweet, C. R., A. A. Ribeiro, and C. R. H. Raetz.** 2004. Oxidation and transamination of the 3″-Position of UDP-N-acetylglucosamine by enzymes from *Acidithiobacillus ferrooxidans*. J. Biol. Chem. **279:**25400-25410. doi: 10.1074/jbc.M400596200.

202. **van Mourik, A., L. Steeghs, J. van Laar, H. D. Meiring, H. J. Hamstra, J. P. van Putten, and M. M. Wosten.** 2010. Altered linkage of hydroxyacyl chains in lipid A of *Campylobacter jejuni* reduces TLR4 activation and antimicrobial resistance. J. Biol. Chem. **285:**15828-15836. doi: 10.1074/jbc.M110.102061.

203. **Heinrichs, D. E., J. A. Yethon, and C. Whitfield.** 1998. Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica*. Mol. Microbiol. **30:**221-232.

204. **Aspinall, G. O., A. G. McDonald, T. S. Raju, H. Pang, A. P. Moran, and J. L. Penner.** 1993. Chemical structures of the core regions of *Campylobacter jejuni* serotypes O:1, O:4, O:23, and O:36 lipopolysaccharides. Eur J Biochem. **213:**1017-1027. doi: 10.1111/j.1432-1033.1993.tb17849.x.

205. **Aspinall, G. O., C. M. Lynch, H. Pang, R. T. Shaver, and A. P. Moran.** 1995. Chemical structures of the core region of *Campylobacter jejuni* O:3 lipopolysaccharide

and an associated polysaccharide. Eur. J. Biochem. **231:**570-578. doi: 10.1111/j.1432-1033.1995.tb20734.x. http://europepmc.org/abstract/MED/7544281; http://dx.doi.org/10.1111/j.1432-1033.1995.0570d.x; http://dx.doi.org/10.1111/j.1432-1033.1995.tb20734.x.

206. **Michael, F. S., C. M. Szymanski, J. Li, K. H. Chan, N. H. Khieu, S. Larocque, W. W. Wakarchuk, J. Brisson, and M. A. Monteiro.** 2002. The structures of the lipooligosaccharide and capsule polysaccharide of *Campylobacter jejuni* genome sequenced strain NCTC 11168. Eur J Biochem. **269:**5119-5136. doi: 10.1046/j.1432-1033.2002.03201.x.

207. **Nam Shin, J. E., S. Ackloo, A. S. Mainkar, M. A. Monteiro, H. Pang, J. L. Penner, and G. O. Aspinall.** 1998. Lipo-oligosaccharides of *Campylobacter jejuni* serotype O:10. Structures of core oligosaccharide regions from a bacterial isolate from a patient with the Miller-Fisher syndrome and from the serotype reference strain. Carbohydr Res. **305:**223-232. doi: http://dx.doi.org/10.1016/S0008-6215(97)00259-0.

208. **Holden, K. M., M. Gilbert, P. J. Coloe, J. Li, and B. N. Fry.** 2012. The role of WlaRG, WlaTB and WlaTC in lipooligosaccharide synthesis by *Campylobacter jejuni* strain 81116. Microb. Pathog. **52:**344-352. doi: 10.1016/j.micpath.2012.03.004. http://europepmc.org/abstract/MED/22445818; http://dx.doi.org/10.1016/j.micpath.2012.03.004.

209. **Poly, F., T. D. Read, Y. Chen, M. A. Monteiro, O. Serichantalergs, P. Pootong, L. Bodhidatta, C. J. Mason, D. Rockabrand, S. Baqar, C. K. Porter, D. Tribble, M. Darsley, and P. Guerry.** 2008. Characterization of two *Campylobacter jejuni* strains for use in volunteer experimental-infection studies. Infect Immun. **76:**5655-5667. doi: 10.1128/IAI.00780-08.

210. **Whitfield, C., N. Kaniuk, and E. Frirdich.** 2003. Molecular insights into the assembly and diversity of the outer core oligosaccharide in lipopolysaccharides from *Escherichia coli* and *Salmonella*. J. Endotoxin Res. **9:**244-249.

211. **Lewis, A. L., N. Desa, E. E. Hansen, Y. A. Knirel, J. I. Gordon, P. Gagneux, V. Nizet, and A. Varki.** 2009. Innovations in host and microbial sialic acid biosynthesis revealed by phylogenomic prediction of nonulosonic acid structure. Proceedings of the National Academy of Sciences. **106:**13552-13557. doi: 10.1073/pnas.0902431106.

212. **Angata, T., and A. Varki.** 2002. Chemical diversity in the sialic acids and related a-keto acids: an evolutionary perspective. Chem. Rev. **102:**439-470. doi: 10.1021/cr000407m. http://dx.doi.org/10.1021/cr000407m.

213. **Zunk, M., and M. J. Kiefel.** 2014. The occurrence and biological significance of the α-keto-sugars pseudaminic acid and legionaminic acid within pathogenic bacteria. Rsc Advances. **4:**3413-3421.

214. **Carlin, A. F., S. Uchiyama, Y. C. Chang, A. L. Lewis, V. Nizet, and A. Varki.** 2009. Molecular mimicry of host sialylated glycans allows a bacterial pathogen to engage

neutrophil Siglec-9 and dampen the innate immune response. Blood. **113:**3333-3336. doi: 10.1182/blood-2008-11-187302 [doi].

215. **Blix, G.** 1936. Über die Kohlenhydratgruppen des Submaxillarismucins. Hoppe-Seyler´ s Zeitschrift Für Physiologische Chemie. **240:**43-54.

216. **Faillard, H.** 1989. The early history of sialic acids. Trends Biochem. Sci. **14:**237-241.

217. **Yu, C., and S. G. Withers.** 2015. Recent developments in enzymatic synthesis of modified sialic acid derivatives. Adv Synth Catal. **357:**1633-1654.

218. **Varki, A.** 2010. Colloquium paper: uniquely human evolution of sialic acid genetics and biology. Proc Natl Acad Sci USA. **107 Suppl 2:**8939-8946. doi: 10.1073/pnas.0914634107 [doi].

219. **Vimr, E., and C. Lichtensteiger.** 2002. To sialylate, or not to sialylate: that is the question. Trends Microbiol. **10:**254-257.

220. **Moran, A. P., M. M. Prendergast, and B. J. Appelmelk.** 1996. Molecular mimicry of host structures by bacterial lipopolysaccharides and its contribution to disease. FEMS Immunol. Med. Microbiol. **16:**105-115.

221. **Vimr, E. R., K. A. Kalivoda, E. L. Deszo, and S. M. Steenbergen.** 2004. Diversity of microbial sialic acid metabolism. Microbiol Mol Biol Rev. **68:**132-153.

222. **Gilbert, M., C. T. Parker, and A. P. Moran.** 2008. *Campylobacter jejuni* lipooligosaccharides; structures and biosynthesis, p. 483-504. *In* I. Nachamkin, C. M. Szymanski, and M. J. Blaser (eds.), *Campylobacter*, 3rd ed., . ASM Press, Washington DC.

223. **van Belkum, A., d. B. van, P. Godschalk, W. Ang, B. Jacobs, M. Gilbert, W. Wakarchuk, H. Verbrugh, and H. Endtz.** 2001. A *Campylobacter jejuni* gene associated with immune-mediated neuropathy. Nat Med. **7:**752-753. http://dx.doi.org/10.1038/89831.

224. **Endtz, H. P., C. W. Ang, N. van den Braak, B. Duim, A. Rigter, L. J. Price, D. L. Woodward, F. G. Rodgers, W. M. Johnson, J. A. Wagenaar, B. C. Jacobs, H. A. Verbrugh, and A. van Belkum.** 2000. Molecular characterization of *Campylobacter jejuni* from patients with Guillain-Barré and Miller Fisher syndromes. J Clin Microbiol. **38:**2297-2301.

225. **Vimr, E., S. Steenbergen, and M. Cieslewicz.** 1995. Biosynthesis of the polysialic acid capsule in *Escherichia coli* K1. J Ind Microbiol Biotechnol. **15:**352-360.

226. **Vann, W. F., D. A. Daines, A. S. Murkin, M. E. Tanner, D. O. Chaffin, C. E. Rubens, J. Vionnet, and R. P. Silver.** 2004. The NeuC protein of *Escherichia coli* K1 is a UDP-N-acetylglucosamine 2-epimerase. J Bacteriol. **186:**706-712.

227. **Vann, W. F., J. J. Tavarez, J. Crowley, E. Vimr, and R. P. Silver.** 1997. Purification and characterization of the *Escherichia coli* Kl *neuB* gene product *N*-acetylneuraminic acid synthetase. Glycobiology. **7:**697-701.

228. **Liu, G., C. Jin, and C. Jin.** 2004. CMP-*N*-acetylneuraminic acid synthetase from *Escherichia coli* K1 is a bifunctional enzyme: identification of minimal catalytic domain for synthetase activity and novel functional domain for platelet-activating factor acetylhydrolase activity. J Biol Chem. **279:**17738-17749. doi: 10.1074/jbc.M400143200 [doi].

229. **Bednarski, M. D., D. C. Crans, R. DiCosimo, E. S. Simon, P. D. Stein, G. M. Whitesides, and M. J. Schneider.** 1988. Synthesis of 3-deoxy-D-manno-2-octulosonate-8-phosphate (KDO-8-P) from D-arabinose: generation of D-arabinose-5-phosphate using hexokinase. Tetrahedron Lett. **29:**427-430.

230. **Linton, D., A. V. Karlyshev, P. G. Hitchen, H. R. Morris, A. Dell, N. A. Gregson, and B. W. Wren.** 2000. Multiple N- acetyl neuraminic acid synthetase (*neuB*) genes in *Campylobacter jejuni*: identification and characterization of the gene involved in sialylation of lipo- oligosaccharide. Mol. Microbiol. **35:**1120-1134. doi: 10.1046/j.1365-2958.2000.01780.x. http:https://dx.doi.org/10.1046/j.1365-2958.2000.01780.x.

231. **Guerry, P., C. P. Ewing, T. E. Hickey, M. M. Prendergast, and A. P. Moran.** 2000. Sialylation of lipooligosaccharide cores affects immunogenicity and serum resistance of *Campylobacter jejuni*. Infection and Immunity. **68:**6656-6662.

232. **Aharoni, A., K. Thieme, C. P. C. Chiu, S. Buchini, L. L. Lairson, H. Chen, N. C. J. Strynadka, W. W. Wakarchuk, and S. G. Withers.** 2006. High-throughput screening methodology for the directed evolution of glycosyltransferases. Nat Meth. **3:**609-614. http://dx.doi.org/10.1038/nmeth899.

233. **Kajihara, Y., S. Akai, T. Nakagawa, R. Sato, T. Ebata, H. Kodama, and K. Sato.** 1999. Enzymatic synthesis of Kdn oligosaccharides by a bacterial α-(2→ 6)-sialyltransferase. Carbohydr. Res. **315:**137-141.

234. **Watson, D. C., W. W. Wakarchuk, S. Leclerc, M. J. Schur, I. C. Schoenhofen, N. M. Young, and M. Gilbert.** 2015. Sialyltransferases with enhanced legionaminic acid transferase activity for the preparation of analogs of sialoglycoconjugates. Glycobiology. **25:**767-773.

235. **Watson, D. C., S. Leclerc, W. W. Wakarchuk, and N. M. Young.** 2010. Enzymatic synthesis and properties of glycoconjugates with legionaminic acid as a replacement for neuraminic acid. Glycobiology. **21:**99-108.

236. **Yu, C., T. Hill, D. H. Kwan, H. Chen, C. Lin, W. Wakarchuk, and S. G. Withers.** 2014. A plate-based high-throughput activity assay for polysialyltransferase from *Neisseria meningitidis*. Anal Biochem. **444:**67-74.

237. **Drouillard, S., T. Mine, H. Kajiwara, T. Yamamoto, and E. Samain.** 2010. Efficient synthesis of 6′-sialyllactose, 6, 6′-disialyllactose, and 6′-KDO-lactose by

metabolically engineered *E. coli* expressing a multifunctional sialyltransferase from the *Photobacterium sp.* JT-ISH-224. Carbohydr Res. **345:**1394-1399.

238. **Audry, M., C. Jeanneau, A. Imberty, A. Harduin-Lepers, P. Delannoy, and C. Breton.** 2010. Current trends in the structure–activity relationships of sialyltransferases. Glycobiology. **21:**716-726.

239. **Cheng, J., S. Huang, H. Yu, Y. Li, K. Lau, and X. Chen.** 2009. Trans-sialidase activity of *Photobacterium damsela* α2, 6-sialyltransferase and its application in the synthesis of sialosides. Glycobiology. **20:**260-268.

240. **Romanow, A., T. G. Keys, K. Stummeyer, F. Freiberger, B. Henrissat, and R. Gerardy-Schahn.** 2014. Dissection of hexosyl- and sialyltransferase domains in the bifunctional capsule polymerases from *Neisseria meningitidis* W and Y defines a new sialyltransferase family. J Biol Chem. **289:**33945-33957. doi: 10.1074/jbc.M114.597773 [doi].

241. **Blixt, O., D. Vasiliu, K. Allin, N. Jacobsen, D. Warnock, N. Razi, J. C. Paulson, S. Bernatchez, M. Gilbert, and W. Wakarchuk.** 2005. Chemoenzymatic synthesis of 2-azidoethyl-ganglio-oligosaccharides GD3, GT3, GM2, GD2, GT2, GM1, and GD1a. Carbohydr. Res. **340:**1963-1972.

242. **Zhang, P., A. J. Zuccolo, W. Li, R. B. Zheng, and C. Ling.** 2009. Probing a sialyltransferase's recognition domain to prepare α (2, 8)-linked oligosialosides and analogs. Chem Comm. 4233-4235.

243. **Pukin, A. V., D. E. Florack, D. Brochu, B. van Lagen, G. M. Visser, T. Wennekes, M. Gilbert, and H. Zuilhof.** 2011. Chemoenzymatic synthesis of biotin-appended analogues of gangliosides GM2, GM1, GD1a and GalNAc-GD1a for solid-phase applications and improved ELISA tests. Org Biomol Chem. **9:**5809-5815.

244. **Jacques, S., J. R. Rich, C. Ling, and D. R. Bundle.** 2006. Chemoenzymatic synthesis of GM3 and GM2 gangliosides containing a truncated ceramide functionalized for glycoconjugate synthesis and solid phase applications. Org Biomol Chem. **4:**142-154.

245. **Morley, T. J., and S. G. Withers.** 2010. Chemoenzymatic synthesis and enzymatic analysis of 8-modified cytidine monophosphate-sialic acid and sialyl lactose derivatives. J. Am. Chem. Soc. **132:**9430-9437.

246. **Mehta, S., M. Gilbert, W. W. Wakarchuk, and D. M. Whitfield.** 2000. Ready access to sialylated oligosaccharide donors. Org. Lett. **2:**751-753.

247. **Olkkola, S., H. . Juntunen, H. Heiska, H. Hyytiainen, and M. Hanninen.** 2010. Mutations in the *rpsL* gene are involved in streptomycin resistance in *Campylobacter coli*. . Microb. Drug Resist. **16:**105. doi: 1089/mdr.2009.0128.

248. **Juntunen, P., S. Olkkola, and M. Hänninen.** 2011. Longitudinal on-farm study of the development of antimicrobial resistance in *Campylobacter coli* from pigs before and

after danofloxacin and tylosin treatments. Vet. Microbiol. **150:**322-330. doi: http://dx.doi.org/10.1016/j.vetmic.2011.02.008.

249. **Juntunen, P., H. Heiska, S. Olkkola, A. -. Myllyniemi, and M. -. Hänninen.** 2010. Antimicrobial resistance in *Campylobacter coli* selected by tylosin treatment at a pig farm. Vet. Microbiol. **146:**90-97. doi: http://dx.doi.org.libproxy.helsinki.fi/10.1016/j.vetmic.2010.04.027.

250. **Lehtopolku, M., P. Kotilainen, M. Haanperä-Heikkinen, U. Nakari, M. Hänninen, P. Huovinen, A. Siitonen, E. Eerola, J. Jalava, and A. J. Hakanen.** 2011. Ribosomal mutations as the main cause of macrolide resistance in *Campylobacter jejuni* and *Campylobacter coli*. Antimicrobial Agents and Chemotherapy. **55:**5939-5941. doi: 10.1128/AAC.00314-11.

251. **Llarena, A. -., C. P. A. Skarp-de Haan, M. Rossi, and M. -. Hänninen.** 2015. Characterization of the *Campylobacter jejuni* population in the barnacle geese reservoir. Zoonoses and Public Health. **62:**209-221. doi: 10.1111/zph.12141.

252. **Feodoroff, B., P. Ellström, H. Hyytiäinen, S. Sarna, M. Hänninen, and H. Rautelin.** 2010. *Campylobacter jejuni* isolates in Finnish patients differ according to the origin of infection. Gut Pathog. **2:**22. doi: 10.1186/1757-4749-2-22. http://dx.doi.org/10.1186/1757-4749-2-22.

253. **Denis, M., C. Soumet, K. Rivoal, G. Ermel, D. Blivet, G. Salvat, and P. Colin.** 1999. Development of a m-PCR assay for simultaneous identification of *Campylobacter jejuni* and *C. coli*. Letters in Applied Microbiology. **29:**406-410.

254. **Chevreux, B., T. Wetter, and S. Suhai.** 1999. Genome sequence assembly using trace signals and additional sequence information. Comput Sci and Biol: Proc Ger Conf on Bioinformatics (GCB). **99:**45-56.

255. **Warren, R., G. Sutton, S. Jones, and R. Holt.** 2007. Assembling millions of short DNA sequences using SSAKE. Bioinformatics. **23:**500-501. http://dx.doi.org/10.1093/bioinformatics/btl629.

256. **Staden, R., K. Beal, and J. Bonfield.** 2000. The Staden package, 1998. Methods Mol Biol. **132:**115-130.

257. **Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A. Pevzner.** 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. **19:**455-477. doi: 10.1089/cmb.2012.0021. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3342519/.

258. **Bolger, A. M., M. Lohse, and B. Usadel.** 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. **30:**2114-2120.

259. **Walker, B. J., T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, and S. K. Young.** 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS One. **9:**e112963.

260. **Aziz, R., D. Bartels, A. Best, M. DeJongh, T. Disz, R. Edwards, K. Formsma, S. Gerdes, E. Glass, M. Kubal, F. Meyer, G. Olsen, R. Olson, A. Osterman, R. Overbeek, L. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, and O. Zagnitko.** 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics. **9:**1-15. doi: 10.1186/1471-2164-9-75. http://dx.doi.org/10.1186/1471-2164-9-75.

261. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. Rajandream, and B. Barrell.** 2000. Artemis: sequence visualization and annotation. Bioinformatics. **16:**944-945. doi: 10.1093/bioinformatics/16.10.944.

262. **The UniProt Consortium.** 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res. **45:**D158-D169. http://dx.doi.org/10.1093/nar/gkw1099.

263. **Mulder, N. J., and R. Apweiler.** 2002. The InterPro database and tools for protein domain analysis*In* Anonymous Current Protocols in Bioinformatics. John Wiley & Sons, Inc.

264. **Seemann, T.** 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics. **30:**2068-2069.

265. **Carver, T., M. Berriman, A. Tivey, C. Patel, U. Böhme, B. G. Barrell, J. Parkhill, and M. Rajandream.** 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics. **24:**2672-2676. doi: 10.1093/bioinformatics/btn529.

266. **McClure, R., D. Balasubramanian, Y. Sun, M. Bobrovskyy, P. Sumby, C. A. Genco, C. K. Vanderpool, and B. Tjaden.** 2013. Computational analysis of bacterial RNA-Seq data. Nucleic Acids Research. **41:**e140-e140. doi: 10.1093/nar/gkt444.

267. **Edgar, R. C., and K. Sjolander.** 2004. A comparison of scoring functions for protein sequence profile alignment*In* Anonymous Bioinformatics.

268. **Edgar, R. C.** 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. **5:**1-19. doi: 10.1186/1471-2105-5-113. http://dx.doi.org/10.1186/1471-2105-5-113.

269. **Abascal, F., R. Zardoya, and M. J. Telford.** 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Research. . doi: 10.1093/nar/gkq291.

270. **Culebro, A., J. Revez, B. Pascoe, Y. Friedmann, M. D. Hitchings, J. Stupak, S. K. Sheppard, J. Li, and M. Rossi.** 2016. Large Sequence Diversity within the Biosynthesis Locus and Common Biochemical Features of Campylobacter coli

Lipooligosaccharides. Journal of Bacteriology. **198:**2829-2840. doi: 10.1128/JB.00347-16.

271. **Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar.** 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution. . doi: 10.1093/molbev/mst197.

272. **Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden.** 2009. BLAST : architecture and applications. BMC Bioinformatics. **10:**421.

273. **Rasko, D. A., G. S. Myers, and J. Ravel.** 2005. Visualization of comparative genomic analyses by BLAST score ratio. BMC Bioinformatics. **6:**2.

274. **Kumar, S., G. Stecher, and K. Tamura.** 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. **33:**1870-1874.

275. **Machado, M. P., B. Ribeiro-Gonçalves, M. Silva, M. Ramirez, and J. A. Carriço.** 2017. Epidemiological surveillance and typing methods to track antibiotic resistant strains using high throughput sequencing, p. 331-356. *In* P. Sass (ed.), Antibiotics: Methods and Protocols. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4939-6634-9_20.

276. **Skarp-de Haan, C., A. Culebro, T. Schott, J. Revez, E. Schweda, M. Hanninen, and M. Rossi.** 2014. Comparative genomics of unintrogressed *Campylobacter coli* clades 2 and 3. BMC Genomics. **15:**129.

277. **Miller, W., E. Yee, K. Jolley, and M. Chapman.** 2014. Use of an improved *atpA* amplification and sequencing method to identify members of the Campylobacteraceae and Helicobacteraceae. Lett Appl Microbiol. **58:**582-590.

278. **Cheng, L., T. R. Connor, J. Sirén, D. M. Aanensen, and J. Corander.** 2013. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. Mol. Biol. Evol. **30:**1224-1228.

279. **Letunic, I., and P. Bork.** 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. **44:**W242-W245.

280. **Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg.** 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protocols. **10:**845-858. http://dx.doi.org/10.1038/nprot.2015.053.

281. **Yates, C. M., I. Filippis, L. A. Kelley, and M. J. E. Sternberg.** 2014. SuSPect: Enhanced prediction of single amino acid variant (SAV) phenotype using network features. J. Mol. Biol. **426:**2692-2701. doi: http://dx.doi.org/10.1016/j.jmb.2014.04.026.

282. **Benkert, P., M. Künzli, and T. Schwede.** 2009. QMEAN server for protein model quality estimation. Nucleic Acids Res. **37:**W510-W514. doi: 10.1093/nar/gkp322. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2703985/.

283. **Darling, A. E., B. Mau, and N. T. Perna.** 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. **5:**e11147. http://dx.doi.org/10.1371%2Fjournal.pone.0011147.

284. **Thomas, M. T., M. Shepherd, R. K. Poole, A. H. M. van Vliet, D. J. Kelly, and B. M. Pearson.** 2011. Two respiratory enzyme systems in *Campylobacter jejuni* NCTC 11168 contribute to growth on L-lactate. Environ. Microbiol. **13:**48-61. doi: 10.1111/j.1462-2920.2010.02307.x.

285. **Olkkola, S., A. Culebro, P. Juntunen, M. Hänninen, and M. Rossi.** 2016. Functional genomics in *Campylobacter coli* identified a novel streptomycin resistance gene located in a hypervariable genomic region. Microbiology. **162:**1157-1166. doi: 10.1099/mic.0.000304.

286. **Guerry, P., C. P. Ewing, T. E. Hickey, M. M. Prendergast, and A. P. Moran.** 2000. Sialylation of lipooligosaccharide cores affects immunogenicity and serum resistance of *Campylobacter jejuni*. Infection and Immunity. **68:**6656-6662. doi: 10.1128/IAI.68.12.6656-6662.2000.

287. **Harrison, L. H., K. A. Shutt, S. E. Schmink, J. W. Marsh, B. H. Harcourt, X. Wang, A. M. Whitney, D. S. Stephens, A. A. Cohn, and N. E. Messonnier.** 2010. Population structure and capsular switching of invasive *Neisseria meningitidis* isolates in the pre-meningococcal conjugate vaccine era—United States, 2000–2005. J. Infect. Dis. **201:**1208-1224.

288. **Harrison, L. H., K. A. Jolley, K. A. Shutt, J. W. Marsh, M. O'Leary, L. T. Sanza, M. C. Maiden, and Maryland Emerging Infections Program.** 2006. Antigenic shift and increased incidence of meningococcal disease. J. Infect. Dis. **193:**1266-1274.

289. **Swartley, J. S., A. A. Marfin, S. Edupuganti, L. J. Liu, P. Cieslak, B. Perkins, J. D. Wenger, and D. S. Stephens.** 1997. Capsule switching of *Neisseria meningitidis*. Proc. Natl. Acad. Sci. U. S. A. **94:**271-276.

290. **Holst Sorensen, M. C., L. B. van Alphen, C. Fodor, S. M. Crowley, B. B. Christensen, C. M. Szymanski, and L. Brondsted.** 2012. Phase variable expression of capsular polysaccharide modifications allows *Campylobacter jejuni* to avoid bacteriophage infection in chickens. Front. Cell. Infect. Microbiol. **2:**11. doi: 10.3389/fcimb.2012.00011 [doi].

291. **Varki, A.** 2006. Nothing in glycobiology makes sense, except in the light of evolution. Cell. **126:**841-845.

292. **Cha, W., R. E. Mosci, S. L. Wengert, C. V. Vargas, S. R. Rust, P. C. Bartlett, D. L. Grooms, and S. D. Manning.** 2017. Comparing the genetic diversity and antimicrobial

resistance profiles of *Campylobacter jejuni* recovered from cattle and humans. Front Microbiol. **8:**.

293. **Boes, J., L. Nersting, E. Nielsen, S. Kranker, C. Enøe, H. Wachmann, and D. L. Baggesen.** 2005. Prevalence and diversity of *Campylobacter jejuni* in pig herds on farms with and without cattle or poultry. J. Food Prot. **68:**722-727.

294. **Lee, H. J., L. L. Lairson, J. R. Rich, E. Lameignere, W. W. Wakarchuk, S. G. Withers, and N. C. J. Strynadka.** 2011. Structural and kinetic analysis of substrate binding to the sialyltransferase Cst-II from *Campylobacter jejuni*. J Biol Chem. **286:**35922-35932. doi: 10.1074/jbc.M111.261172.

295. **Sundaram, A., L. Pitts, K. Muhammad, J. Wu, M. Betenbaugh, R. Woodard, and W. Vann.** 2004. Characterization of *N*-acetylneuraminic acid synthase isoenzyme 1 from *Campylobacter jejuni*. Biochem. J. **383:**83. http://www.biochemj.org/content/383/1/83.abstract.